Layered Space-Time Structure for MIMO-OFDM Systems

A Thesis
Presented to
The Academic Faculty

by

Jianxuan Du

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
August 2005

Layered Space-Time Structure for MIMO-OFDM Systems

Approved by:

Professor Ye (Geoffrey) Li, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Douglas Williams
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Mary Ann Ingram
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Greg Durgin
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Doron Lubinsky
School of Mathematics
Georgia Institute of Technology

Date Approved: July 12, 2005

To my loving Dad and Mom,

for their understanding and support through the years.

ACKNOWLEDGEMENTS

I would like to extend my thanks to my adviser Ye (Geoffrey) Li, for guiding me through the years of Ph.D. study. His insightful remarks and encouragement have been of great help to my research as well as to my future career. But for his knowledge and experience in research coming to help in time of need, my Ph.D. study would not have been so smooth and enjoyable. I believe what I have learned during the past few years, both in research and in attitude toward life, has become an invaluable part of my personality that leads me to future success.

I would also like to express my appreciation to Dr. Douglas Williams, Dr. Mary Ann Ingram, Dr. Greg Durgin, Dr. Doron Lubinsky for being on my defense committee. Their remarks and comments have made significant contributions to the improvement of my work. In addition, inspiring and spirited talks with other members of Information Transmission and Processing Laboratory (ITP), Guocong Song, Taewon Hwang, Jingnong Yang, Hua Zhang, Uzoma Anaso Onunkwo, and Ghurumuruhan Ganesan have always turn out to be very rewarding. Exchange of ideas and views in different perspectives helps me keep an open and creative mind.

Last, I wish to give thanks to my parents for giving me a clear view of my career path and for all their persistent support and understanding during the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

By employing multiple antennas at the transmitter and receiver of a communication system, transmit and receive diversity can be exploited to combat fading of wireless channels. The pioneering studies by Telatar, Foschini, and Gans [1], [2] show a huge capacity increase by employing multiple antennas in fading environments. *Multiple-input multiple-output* (MIMO) techniques have attracted significant attention ever since.

Various approaches have been tried to achieve this huge capacity gain. *Space-time trellis codes* (STTC) are proposed in [3] together with construction criteria. *Space-time block codes* (STBC) such as Alamouti's code [4] based on orthogonal designs, are also attractive for their simplicity of detection. A third popular approach, layered structure, is based on signal processing and conventional one-dimensional coding, which includes vertical BLAST (*Bell Laboratories Layered Space Time*) and diagonal BLAST [5], [6], [7]. In layered systems, the input data stream is demultiplexed, independently coded using one-dimensional coding, and sent via different transmit antennas simultaneously. The received signal from each substream is separated by nulling according to *zero-forcing* (ZF) or *minimum mean square-error* (MMSE) criterion and *successive interference cancellation* (SIC).

The low complexity of layered processing makes the layered structure a promising candidate for MIMO systems with a large number of transmit antennas and higher order modulation. Another advantage is the backward compatibility of layered structure, where industry standard one-dimensional coding such as convolutional coding can be used for each layer, thus providing smooth transition from *single-input single-output* (SISO) to MIMO systems.

For broadband systems, *orthogonal frequency division multiplexing* (OFDM) appears promising for its immunity against delay spread. In addition, OFDM is especially suitable for frequency-selective MIMO systems since the introduction of orthogonal subcarriers makes system design and implementation as simple as those for flat-fading channels. The proposed research is focused on the layered structure for MIMO-OFDM systems, where several techniques are proposed for performance enhancement, namely, channel estimation based on subspace tracking, parallel detection of group-wise space-time codes by predictive soft interference cancellation, quasi-block diagonal *low-density parity-check codes* (LDPC) coding and statistical data rate allocation for layered systems.

For coherent detection, channel parameters are needed and channel estimation algorithm plays an important role in determining system performance. For MIMO-OFDM systems, rank reduction by some linear transform matrix is necessary for channel estimation with a single OFDM block. In the proposed research, we propose a channel estimation algorithm for MIMO-OFDM systems, which has considerably less leakage than DFT-based channel estimators. This estimation algorithm uses the optimum low-rank channel approximation obtained by tracking the frequency autocorrelation matrix of the channel response.

Then parallel detection algorithm is proposed for a generalized layered system with group-wise space-time coding, where the structure of particular component space-time code trellises is exploited using partial information from the Viterbi decoder of the simultaneously decoded interfering component codes.

Next we incorporate the layered structure with LDPC to develop a quasi-block diagonal LDPC space-time structure. The lower triangular structure of the parity check matrix introduces correlation between layers. Each layer, as a part of the whole codeword, can be decoded while taking information from other undetected layers to improve the decoding performance.

In the end, a modified layered structure is proposed where the layer detection order is fixed and the data rate for each layer is allocated based on the detection order and channel statistics. With Gaussian approximation of layer capacities, we derive the optimum data rate allocation.

CHAPTER I

INTRODUCTION

Channel fading in wireless environments has always been a big challenge for engineers before high-speed data transmission can be realized. Spatial diversity introduced by deploying multiple antennas at the transmitter and/or the receiver is an effective technique to combat fading effect. Ever since the introduction of the idea of space-time coding [3] and information-theoretic analysis of MIMO systems in fading channels [1], MIMO systems have become an attractive technique that dramatically increases channel capacity and thus enhances spectral efficiency.

1.1 MIMO Wireless Channel

1.1.1 Wireless Channel

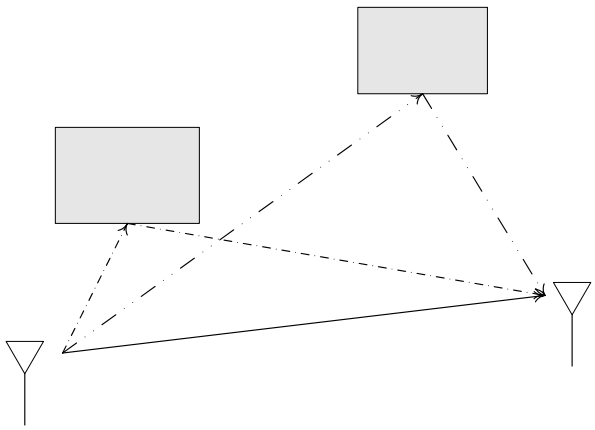


Figure 1: Multipath propagation of signals in a wireless channel.

In a wireless system, the radio signal from the transmitter travels through different

propagation paths to reach the receiver, as shown in Figure 1. The superposition of multipath signals can add up constructively or destructively, which causes channel fading, i.e., significant fluctuation of received signal energy.

The complex baseband equivalent of a fading channel impulse response can be expressed as,

$$h(t,\tau) = \sum_{l=1}^{L} \alpha_l(t)\,\Delta(\tau - \tau_l), \tag{1}$$

where $\alpha_l(t)$'s are wide-sense stationary narrow-band complex Gaussian processes and are assumed to be independent among different paths. L is the number of multipaths. $\Delta(t)$ is the delta function and τ_l is the relative delay of the l-th multipath. In addition, we assume the same normalized time-domain correlation function for all paths, i.e.,

$$E\{\alpha_l(t+\Delta t)\,\alpha_m^*(t)\} = \begin{cases} \sigma_l^2 r_t(\Delta t), & l = m, \\ 0, & l \neq m. \end{cases} \tag{2}$$

Without loss of generality, we assume that the total average power of the channel impulse response to be unity, *i.e.*,

$$\sum_l \sigma_l^2 = 1. \tag{3}$$

1.1.2 MIMO Channel

MIMO systems formed by multiple transmit and receive antennas are under intense research recently for its attractive potential to offer great capacity increase. MIMO transmission capitalizes on the fact that signals at different antennas experience independent fading in a scattering environment, if the antennas are well separated. For a flat-fading MIMO system with N_t transmit and N_r receive antennas, the received signal at the j-th receive antenna can be expressed as:

$$r_j = \sum_{i=1}^{N_t} H_{ji} s_i + w_j, \tag{4}$$

where s_i is the symbol transmitted from the i-th transmit antenna, H_{ji} is the channel impulse response corresponding to the propagation path between the i-th transmit

and the j-th receive antenna, and, $w_j[n,k]$ is additive (complex) Gaussian noise that is assumed to be *independent and identically distributed* (i.i.d.) with zero-mean and variance ρ. A system diagram for a flat-fading MIMO system is shown in Figure 2. Note that the transmitted signals from all transmit antennas overlap in time, space and frequency so that at the receiver, the received signal is a superposition of all transmitted signals distorted by the channel plus noise.

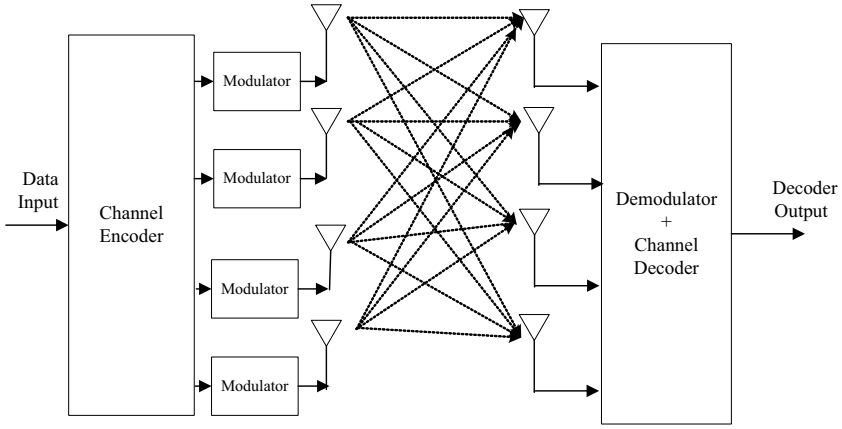


Figure 2: System diagram of a MIMO system with four transmit and four receive antennas.

It has been proved in [1] that in a system where the instantaneous channel state information is not known at the transmitter, the instantaneous open-loop channel capacity is

$$C = \log\left[\det\left(\mathbf{I}_{N_r} + \frac{1}{N_t\rho}\mathbf{H}\mathbf{H}^H\right)\right], \tag{5}$$

where $(\mathbf{H})_{ji} = H_{ji}$.

We consider the case where channels between different receive antennas are statistically independent, then the ergodic capacity is

$$C_{erg} = E\left\{\log\left[\det\left(\mathbf{I}_{N_r} + \frac{1}{N_t\rho}\mathbf{H}_w\mathbf{H}_w^H\right)\right]\right\},$$

$\mathbf{H}_w$ is an $N_r \times N_t$ random matrix with i.i.d. $CN(0,1)$ entries. In addition, information-theoretic analysis shows that the ergodic capacity increases linearly with the number

of antenna elements [2] [1].

1.2 OFDM Systems

OFDM is a multicarrier transmission technique proposed in the mid-1960's. In a frequency-selective channel, the delay spread of the channel impulse response introduces inter-symbol interference (ISI) in a single carrier system, which causes severe system performance degradation if not taken care of. OFDM effectively counters the channel delay spread by converting the channel into a number of overlapping but mutually orthogonal subchannels in the frequency domain. The spectrum of an OFDM signal is shown in Figure 3 as an example. The main lobe of each subcarrier lies on the nulls of all other subcarriers so that there is no mutual interference.

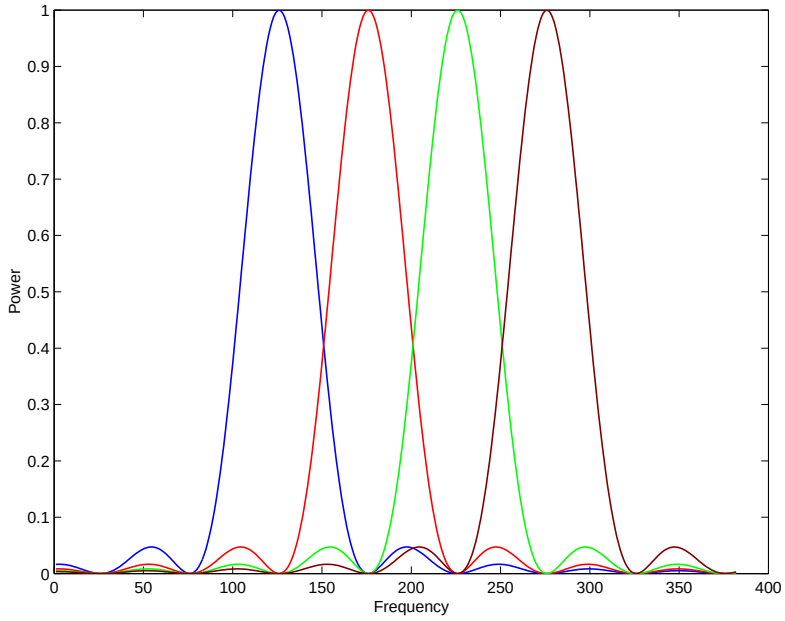


Figure 3: Power spectrum of subcarriers in an OFDM system .

Though it has been proposed for quite some time, OFDM was recently widely recognized as an ideal method for high-speed wireless data transmission, mainly due to the availability of high-speed implementation of FFT by modern VLSI design. OFDM

has been successfully adopted in many systems such as Digital Video Broadcasting-Terrestrial (DVB-T), Asymmetric Digital Subscriber Line (ADSL), and in wireless LAN standards such as IEEE 802.11a and IEEE 802.11g. OFDM is also being considered as a promising candidate for future generation high-speed digital communication systems such as IEEE 802.11n.

The relationship actual transmitted signal with the complex baseband signal is as follows.

$$s(t) = Re\left(s_{cb}(t)e^{j\omega_c t}\right).$$

where $Re(.)$ denotes taking real part of a complex variable, and ω_c is the carrier center frequency. In an OFDM system, the data are modulated onto different subcarriers and sent in blocks. The complex baseband representation of the modulated signal during the n-th OFDM block is

$$s_{cb}(t) = \begin{cases} \sum_{k=0}^{K-1} s[n,\ k]e^{-j2\pi k\Delta f(t-T_{FFT}-nT_s)} & nT_s - T_{GI} \le t < nT_s, \\ \sum_{k=0}^{K-1} s[n,\ k]e^{-j2\pi k\Delta f\omega(t-nT_s)} & nT_s \le t < nT_s + T_{FFT}. \end{cases} \tag{6}$$

where s[k] is the signal to be transmitted at the k-th subcarrier, $\Delta f = \frac{B}{K}$ is the subcarrier spacing, B is the total bandwidth, and K is the total number of subcarriers, $T_{FFT} = \frac{1}{\Delta f}$ is the IFFT/FFT period and T_{GI} is the guard interval. Hence the total duration of an OFDM block is $T_s = T_{FFT} + T_{GI}$. Usually the multicarrier modulation is implemented using modern digital signal processing techniques such as IFFT to avoid a large number of modulators/demodulator and filters.

From Equation (6), it can be seen that during the guard interval $[nT_s - T_{GI},\ nT_s)$, the copy of $s_{cb}(t)$ at the interval $[nT_s + T_{FFT} - T_{GI}, nT_s + T_{FFT})$ is inserted, called cyclic prefix. The cyclic prefix is inserted to maintain the orthogonality between signals at different subcarriers and consecutive OFDM blocks. With the insertion of cyclic prefix, the received signal is complete free of ISI during $0 \le t < T_{FFT}$, if the maximum delay of the channel impulse response is less than T_{GI}. Thus, at the

receiver, the signals at different subcarriers can be separated by FFT of the received signal after the part during the guard interval is discarded.

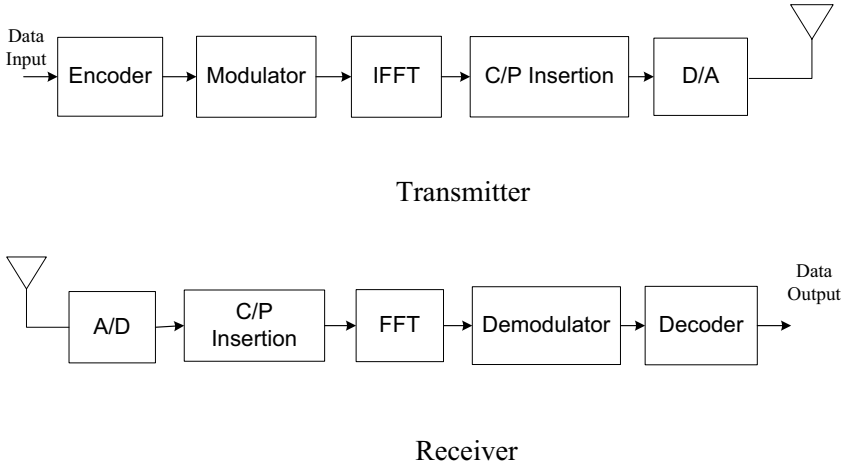


Figure 4: System diagram of an OFDM system.

The insertion of cyclic prefix converts the linear convolution of the channel impulse response with the transmitted signal to cyclic convolution. Therefore, assuming the channel response stays constant during each OFDM block, the channel frequency response for the kth subcarrier during the n-th OFDM block is

$$H\left[n,k\right]=\sum_{l}\alpha_{l}\left(nT_{s}\right)e^{-j2\pi k\Delta f\tau_{l}}, \tag{7}$$

1.3 MIMO-OFDM Systems

For a MIMO-OFDM system with N_t transmit and N_r receive antennas, the received signal at the k-th subcarrier of the n-th block from the j-th receive antenna can be expressed as:

$$r_{j}\left[n,k\right]=\sum_{i=1}^{N_{t}}H_{ji}\left[n,k\right]s_{i}\left[n,k\right]+w_{j}\left[n,k\right], \tag{8}$$

for $j=1,\ \ldots,\ N_r$ and $k=0,\ \ldots,\ K-1$, where $s_i\left[n,k\right]$ is the symbol transmitted from the i-th transmit antenna at the k-th subcarrier of the n-th block, $H_{ji}\left[n,k\right]$ is the

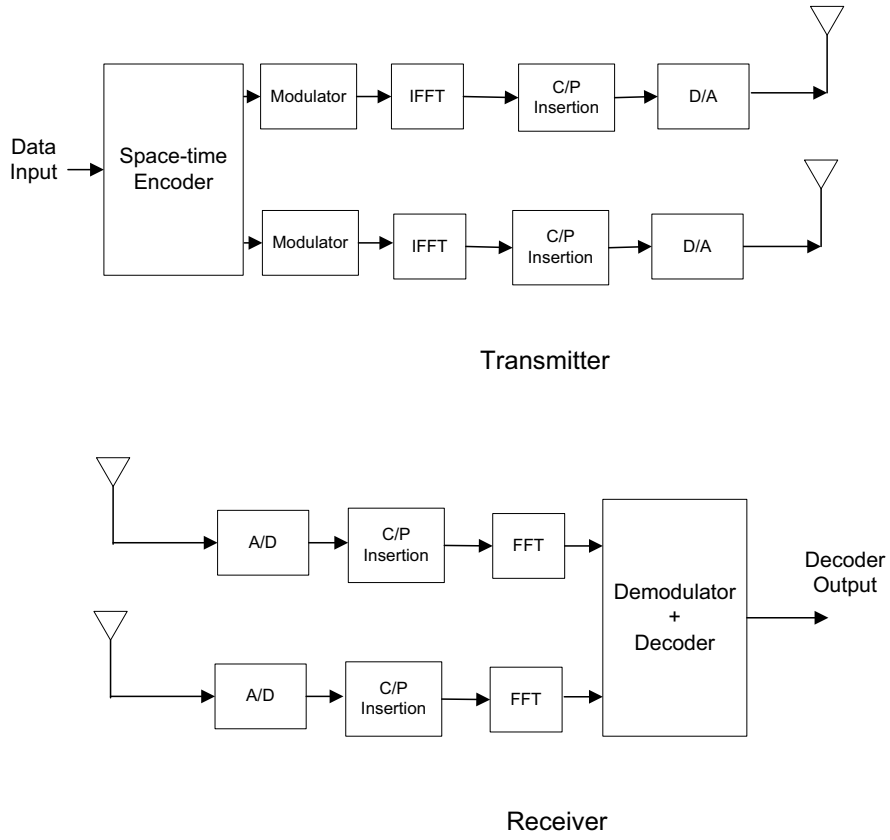


Figure 5: System diagram of a MIMO-OFDM system with two transmit and two receive antennas.

channel's frequency response at the k-th subcarrier of the n-th block corresponding to the i-th transmit and the j-th receive antenna, and, $w_j\,[n,k]$ is additive (complex) Gaussian noise that is assumed to be i.i.d. with zero-mean and variance ρ.

The instantaneous open-loop channel capacity is

$$C = \frac{1}{K}\sum_{k=0}^{K-1} C_k,$$

where C_k is the instantaneous open-loop capacity of the channel at the k-th subcarrier as expressed by Equation (5).

The ergodic capacity for frequency-selective MIMO-OFDM systems is the same

as that of flat-fading systems [1]. However, due to the additional diversity introduced by frequency selectivity, frequency-selective MIMO-OFDM systems should provide stabler performance if the frequency diversity is properly utilized. This improvement is reflected by outage capacity, which is defined as C_{out} such that $Pr\{C < C_{out}\} = p_{out}$, where p_{out} is the outage probability. For the flat-fading case, both capacities can be expressed in terms of integrals involving Laguerre polynomials [1]. For outage capacity of frequency-selective channels, the analysis is even more complicated.

The ergodic capacities (in bps/Hz) of 2×2, 4×2, and 4×4 systems are shown in Figure 6. The capacity of a 4×2 is close to that of a 2×2 system and there is a considerable capacity increase for a 4×4 system, especially at high *signal-to-noise ratio*'s (SNR). This is due to the fact that system capacity is mainly linear in the minimum of the number of transmit antennas and the number of receive antennas. Figure 7 gives the outage capacities of a 2×2 flat-fading channel with outage probability 10%, 1%, and 0.1%, respectively. Figure 8 gives the corresponding outage capacities for OFDM sytems with 20MHz and 256 subcarriers. The bandwidth penalty due to cyclic prefix is neglected. It is clear that the presence of frequency diversity due to channel delay spread yields a more consistent performance and outage capacities for different outage probabilities are very close.

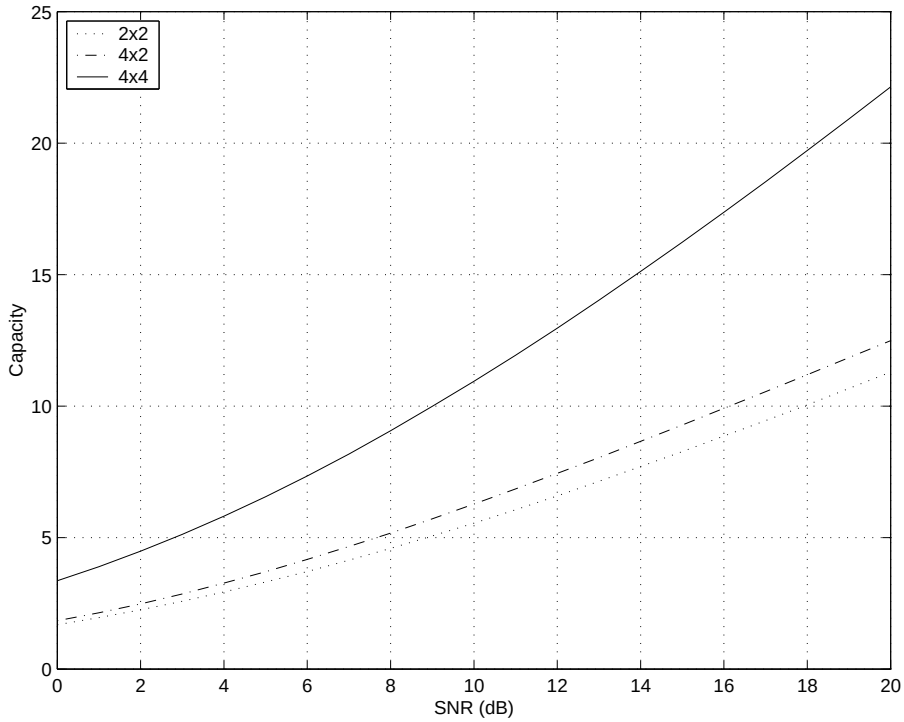


Figure 6: Ergodic capacity of Fading channels.

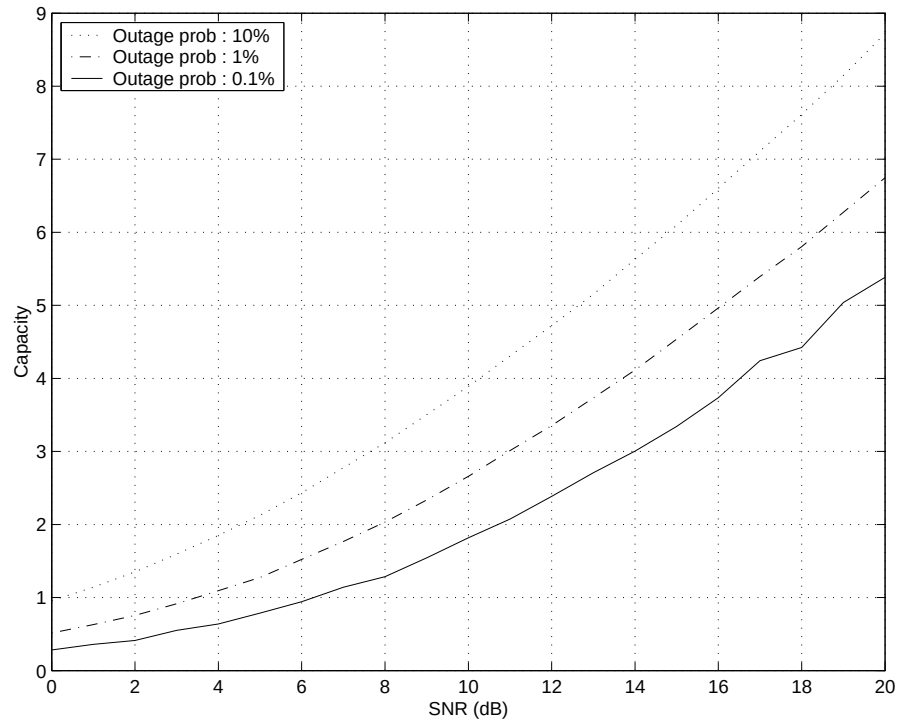


Figure 7: Outage capacity of 2x2 flat-fading channels.

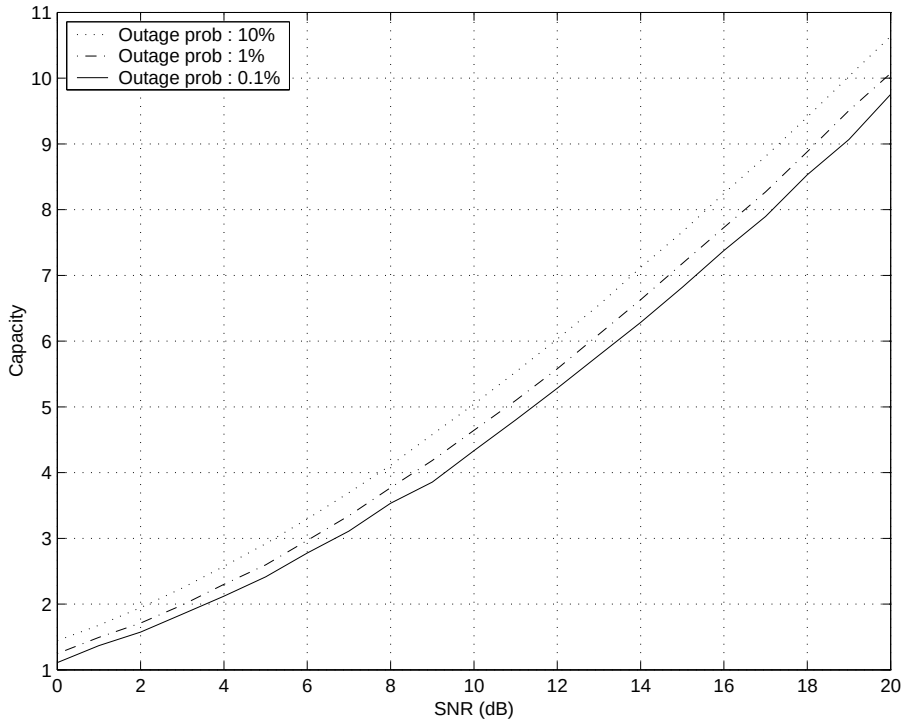


Figure 8: Outage capacity of 2x2 OFDM systems, TU delay profile .

CHAPTER II

LITERATURE REVIEW

We have thus far discussed the capacity of MIMO-OFDM systems. Next we will summarize various techniques proposed in an attempt to achieve this capacity and give their respective advantages and disadvantages.

2.1 Space-Time Block Coding

Alamouti's code [4], one special case of space-time block codes, is probably the simplest form of space-time structure. The orthogonal structure of the transmitted signal decouples the two symbols sent at the same time and the maximum likelihood detection of the 2-D structure is greatly simplified. It has been shown that Alamouti's code gains full diversity for flat-fading channels and achieves system capacity in a 2×1 system. However, there are several factors that limit the use of space-time codes from orthogonal designs.

- *Existence.* It is shown in [8] that no full rate orthogonal space-time block codes exist for complex constellations when there are more than two transmit antennas.

- *Capacity loss.* Even for Alamouti's code, it has been proved that for more than one receive antenna, the code does not achieve the channel capacity, and the capacity loss is considerable at high SNR's [9].

STBC can be directly used in space-frequency domain, as in [10]. However, since the block code is virtually designed by repetition of the same symbol, it requires the channel response stays the same, which implies STBC can only be used on two

consecutive subcarriers, if feasible at all, and itself can not make use of the frequency diversity. Therefore, additional outer coding is needed to gain the frequency diversity in OFDM systems.

2.2 Space-Time Trellis Coding

After the pioneering work of Tarokh [3], space-time coding has been seen as an effective way to achieve the huge potential channel capacity of MIMO systems. The idea of space-time coding is to treat the transmitted signal vector as a *symbol* in a codeword. By investigating pairwise probability of error and computing an upper bound on the average probability of error, the rank criterion and determinant criterion are proposed in [3] in an effort to maximize diversity gain and coding gain, respectively.

STTC has been extensively studied by numerous researchers. [11] obtains optimum space-time convolutional codes by exclusive search of the generator matrix. Seshadri proposes to use a super set of orthogonal space-time block codes as *symbols* in a codeword, and use set partitioning for trellis code design. Siwamogsatham [12] uses unitary transform to expand the STBC code set and also resorts to set partitioning for trellis design. Existent codes designed for flat-fading channels are applied to OFDM in [13], and the design of space-time-frequency trellis codes that exploits the additional frequency diversity is studied in [14] and [15].

The major challenge for STTC is that the *maximum-likelihood* (ML) decoding complexity grows exponentially with the number of transmit antennas. As more transmit antennas are used, the complexity could become a big issue for high-speed real time decoding. Another issue is that the code design becomes more complicated, even with the given criteria, the search for the optimum code is still an open problem.

2.3 Layered Structure

To simplify space-time structure, Foschini [6] [5] proposes to form layers where streams of data are independently coded and decoded. Linear interference suppression(nulling) accompanied by decision-directed interference cancellation is used for detection of each stream (layer). To detect the layered space-time structure, the number of receive antennas should be greater or equal to the number of transmit antennas. The layered structure is generalized in [7].

Two layered structures are popular research topics. V-BLAST associates each stream to a fixed transmit antenna whereas D-BLAST cyclically changes the association and codewords are dispersed across diagonals in space-time. The readily available *forward-error-correction* (FEC) one-dimensional coding can be used for each layer, and the data rate can be easily adjusted by just changing the cod rate. However, due to the nature of the detection algorithm, powerful FEC codes are needed to reduce the effect of error propagation. This is more of a problem for V-BLAST since there is less diversity for each layer. For MIMO-OFDM, frequency diversity is exploited by the FEC code that spans all significantly different subcarriers. The performance of V-BLAST-OFDM is studied in [16].

As argued in [9], the reason for the capacity loss of orthogonal designs lies in the lack of *degrees of freedom*: many observed signals correspond to only a few repeated symbols. Thus it is proposed in [9] that each symbol is linearly dispersed to all transmit antennas and all times. The paper sets up a general framework that includes V-BLAST and orthogonal codes as special cases. The dispersion matrices are found by maximizing the mutual information.

2.4 Group-wise Space-Time Coding (GSTC): Combined Processing

Given the advantages and disadvantages of STTC and layered structure, group-wise space-time coding is proposed as a tradeoff. Another motivation is that we can apply existent space-time codes to a higher dimension, especially for STBC. The main idea is to divide the set of transmit antennas into several subsets. Each subset corresponds to the output of an independent space-time encoder, which forms a subcode. Similar to layered structure, interference between different space-time subcodes is suppressed by linear processing. Group-wise space-time coding may thus be treated as a generalization of the layered structure.

Systems with four transmit antennas using two-space-time codes are proposed in [17] for flat-fading channels and for MIMO-OFDM systems in [18]. Partial orthogonal space-time block codes are proposed in [19] for systems with four transmit antennas.

With the comparison above, layered structure is a promising technology for low-complexity high-speed wireless communication systems. However, there is still a considerable performance gap between a practical V-BLAST system and the Shannon capacity. Therefore, we dedicate our research effort to developing techniques that improve upon the conventional layered structure.

CHAPTER III

CHANNEL ESTIMATION BASED ON SUBSPACE TRACKING

Channel parameters are needed to coherently decode the transmitted signal. *Least square* (LS) channel estimation for MIMO-OFDM systems has been addressed in [20] [21]. But if the multipaths are not sample-spaced, the well-known leakage problem for DFT-based channel estimation [22] induces an error floor for estimation error. To reduce this error floor, more taps have to be used, which not only increases computational complexity but also makes estimation problem more ill-conditioned and thus enhances noise. As an alternative, channel estimation algorithm based on parametric model has been proposed in [23] and extended to MIMO-OFDM in [24]. By using *estimation of signal parameters via rotational invariance techniques* (ESPRIT), channel estimation is achieved by only estimating a few coefficients corresponding to different multipath delays to yield smaller error. However, when the channel varies slowly, the number of training symbols needed for initial multipath delay acquisition may become too large to be practical. Another parametric model-based algorithm needs exhaustive search to find the near-maximum likelihood time of arrival estimation [25], which suffers high computational complexity.

We propose to estimate the dominant subspace spanned by the frequency responses of different multipaths. Due to the fact that in a mobile environment the multipath delays vary slowly the subspace can be tracked by using training OFDM blocks sent at regular interval. The coefficients corresponding to different eigenvectors of the resultant subspace, which vary much faster, are updated every OFDM block

based on decision-directed LS channel estimation and further refined by a robust filter [26] to exploit time domain correlation.

3.1 Decision-Directed Least-Square Channel Estimation

Decision-directed channel estimation is a good option for saving overhead of using dedicated training/pilot symbols or tracking fast-varying channels. After detection of each OFDM block, we use the detected symbols for channel estimation. Since channel responses are independent among different transmit-receive antenna pairs, we consider the channel estimation for one particular receive antenna and omit the receive antenna subscript j in equation (8) to get:

$$r[n,k] = \sum_{i=1}^{N_t} H_i[n,k]\, s_i[n,k] + w[n,k]. \tag{9}$$

Define

$$\mathbf{r}[n] \triangleq (r[n,0],\ r[n,1],\ \cdots,\ r[n,K-1])^T,$$

$$\hat{\mathbf{D}}_i[n] \triangleq diag\{\hat{s}_i[n,0],\ \cdots,\ \hat{s}_i[n,K-1]\},$$

$$\mathbf{w}[n] \triangleq (w[n,0],\ w[n,1],\ \cdots,\ w[n,K-1])^T,$$

where $\hat{s}_i[n,k]$'s are detected symbols. Let rank-χ approximation of the channel response be:

$$(H_i[n,0],\ \cdots,\ H_i[n,K-1])^T = \mathbf{U}\mathbf{h}_i[n], \tag{10}$$

where

$$\mathbf{h}_i[n] \triangleq (h_i[n,0],\ \cdots,\ h_i[n,\chi-1])^T,$$

and $\mathbf{U}$ is the rank-χ transform matrix. For DFT-based channel estimator as in [20],

it is just

$$\mathbf{U} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_K & \cdots & W_K^{\chi-1} \\ \vdots & \cdots & \ddots & \vdots \\ 1 & W_K^{(K-1)} & \cdots & W_K^{(\chi-1)(K-1)} \end{pmatrix},$$

which approximates the channel response as

$$H_i[n,k] = \sum_{l=0}^{\chi-1} h_i[n,l]\, W_K^{kl}, \tag{11}$$

where $W_K = e^{-j\frac{2\pi}{K}}$. Thus, with correct detection, equation (9) can be written in matrix notation as

$$\mathbf{r}[n] = \sum_{i=1}^{N_t} \hat{\mathbf{D}}_i[n]\, \mathbf{U}\mathbf{h}_i[n] + \mathbf{w}[n]. \tag{12}$$

Let

$$\hat{\mathbf{h}}[n] \triangleq \left(\hat{\mathbf{h}}_1^T[n],\ \hat{\mathbf{h}}_2^T[n],\ \cdots,\ \hat{\mathbf{h}}_{N_t}^T[n]\right)^T,$$

from [20], the LS estimate of the channel is

$$\hat{\mathbf{h}}[n] = \left(\mathbf{T}^H[n]\,\mathbf{T}[n]\right)^{-1}\mathbf{T}^H[n]\,\mathbf{x}[n] \tag{13}$$

where

$$\mathbf{T}[n] = \left(\hat{\mathbf{D}}_1[n]\,\mathbf{U},\ \hat{\mathbf{D}}_2[n]\,\mathbf{U},\ \cdots,\ \hat{\mathbf{D}}_{N_t}[n]\,\mathbf{U}\right). \tag{14}$$

In [20], it has been shown that the estimate is unbiased and the mean-square error of the channel estimator is

$$\begin{aligned} MSE[n] &\triangleq \frac{1}{N_t\chi} E\left\{\left\|\hat{\mathbf{h}}[n] - \mathbf{h}[n]\right\|^2\right\} \\ &= \frac{1}{N_t\chi}\mathbf{Tr}\left\{\rho\left(\mathbf{T}^H[n]\,\mathbf{T}[n]\right)^{-1}\right\}. \end{aligned} \tag{15}$$

It can be proved in Appendix A using the bordering theorem for Hermitian matrices [27] that condition number of $\mathbf{T}^H[n]\,\mathbf{T}[n]$ increases with χ. It implies that the channel estimation becomes more ill-conditioned as the number of parameters to be estimated

increases. Thus we should choose the number of parameters as small as possible while preserving energy of the channel response, which is the reason for tracking the optimum basis to form the transform matrix $\mathbf{U}$ in the next section .

3.2 Subspace Tracking for Low-Rank Representation of Channel Responses

The major problem of decision-directed channel estimation is the randomness of the symbol sequences during data transmission mode. For example, when the symbol sequences from any two of the transmit antennas are same or close, it is impossible or hard to distinguish channel responses corresponding to different transmit antennas. The more the number of transmit antennas, the more likely the channel is unidentifiable, or the more ill-conditioned channel identification is. Furthermore, to reduce the leakage of decision-directed DFT-based channel estimation in MIMO-OFDM systems, the number of taps representing the channel frequency response has to be increased, which will make channel identification more ill-conditioned at the same time, as shown in the previous subsection. Moreover, increasing the number of taps makes the inverse operation of matrices in Equation (22) more complicated. Hence, it is essential for low-complexity and high-performance channel estimator to reduce the number of parameters to be estimated while preserving most energy of channel frequency responses during the data transmission mode. Therefore, we will develop subspace tracking approaches to estimate channel parameters.

Let the $K \times K$ channel autocorrelation matrix $(\mathbf{R}_f)_{k_1,k_2} = E\{H[k_1] H^*[k_2]\}$ have singular value decomposition as:

$$\mathbf{R}_f = \mathbf{U}_f \Lambda \mathbf{U}_f^H \tag{16}$$

where $\mathbf{U}_f$ is a $K \times K$ unitary matrix. $\Lambda = diag\{\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_K\}$, $\lambda_1 \geq \lambda_2 \geq \cdots > 0$. From [28], optimum rank-χ estimator is to select eigenvectors $\mathbf{u}_1,\ \mathbf{u}_2,\ \cdots,\ \mathbf{u}_\chi$ corresponding to the χ biggest eigenvalues. Then the optimum rank-χ transform

matrix is:

$$\mathbf{U}_{opt} = (\mathbf{u}_1,\ \mathbf{u}_2,\ \cdots,\ \mathbf{u}_\chi)\,. \tag{17}$$

Therefore, channel autocorrelation matrix is needed here for the optimum low-rank channel estimation.

Since channel autocorrelation is determined by channel delay profile, which varies slowly, periodically sent training blocks can be used for tracking channel autocorrelation as well as for channel estimation. In [21] and [29] optimal training sequences have been proposed to maximally separate frequency responses of different transmit antennas while preserving most energy of each channel response. The training sequences are

$$s_i\,[n,k] = s_1\,[n,k]\,W_K^{-K_0(i-1)k},$$

for i=1, 2, ..., N_t, where $K_0 = \lfloor K/N_t \rfloor \geq \lceil t_d/t_s \rceil$, t_d is the maximum delay spread, and t_s is the sampling interval that is equal to $\frac{1}{K\Delta f}$. It is known that the leakage introduced by the DFT-based approximation is decreased as K_0 increases. Since we are using well designed training sequences rather than arbitrary data sequences, we can set K_0 to be big enough such that the leakage is negligible while introducing little aliasing between different channel responses [21]. The procedure to separate channel responses can be described in Table 3.2.

The dimension χ of the subspace can be either determined by *minimum description length* (MDL) criterion [30] that is not accurate for low SNR or slow channel variation, or by the approach in [28] and [31]. We just choose the latter approach for its simplicity and effectiveness; therefore, $\chi = \lceil t_d/t_s \rceil$. Subspace tracking approach can be summarized Table 2, which is modified LORAF 1 in [32].

It should be noted that the robust channel estimator depends only on the subspace spanned by the dominant eigenvectors rather than the particular eigenvectors. Let

$$\hat{\mathbf{U}}_i\,[n] = \mathbf{U}_{opt}\mathbf{Q}_i\,[n]\,, \tag{18}$$

Table 1: Channel separation using the optimum training sequences

a) During each training block, $\eta[n,k] = r[n,k] \cdot s_1^*[n,k]$;

s) Perform IFFT on $(\eta[n,0],\ \eta[n,1],\ \cdots,\ \eta[n,K-1])$ to get $(\zeta[n,0],\ \zeta[n,1],\ \cdots,\ \zeta[n,K-1])$;

c) For the channel response of transmit antenna i, circularly left shift $(\zeta[n,0],\ \zeta[n,1],\ \cdots,\ \zeta[n,K-1])$ by $(i-1)K_0$ to get $(\zeta'[n,0],\ \zeta'[n,1],\ \cdots,\ \zeta'[n,K-1])$.
Let $\zeta''[n,k] \triangleq \begin{cases} \zeta'[n,k] & k \in [0,\ K_0-1], \\ 0 & \text{else;} \end{cases}$

d) A channel estimate $\hat{H}_i[n,\ k]$ is obtained by performing FFT on $(\zeta''[n,0],\ \zeta''[n,1],\ \cdots,\ \zeta''[n,K-1])$.

Table 2: Subspace tracking for channel estimation

Initialization:
$(\mathbf{U}_i[0])_{k,l} = W_K^{kl}/\sqrt{K},\ 0 \le k \le K-1, 0 \le l \le \chi-1;$
$\Phi[0] = \mathbf{I},\ 0 \le \alpha \le 1;$

During each training block:
input $\mathbf{v}_i[n] = \left(\hat{H}_i[n,\ 0],\ \hat{H}_i[n,\ 1],\ \cdots,\ \hat{H}_i[n,\ K-1]\right)^T$
$\mathbf{c}_i[n] = \mathbf{U}_i^H[n-1]\,\mathbf{v}_i[n]$
$\mathbf{A}_i[n] = \alpha\mathbf{A}_i[n-1]\,\Phi_i[n-1] + (1-\alpha)\mathbf{v}_i[n]\,\mathbf{c}_i^H[n]$
$\mathbf{A}_i[n] = \mathbf{U}_i[n]\,\mathbf{R}_i[n]$
$\Phi_i[n] = \mathbf{U}_i^H[n-1]\,\mathbf{U}_i[n]$
Low rank channel approximation:
$\mathbf{v}_i[n] = \mathbf{U}_i[n]\,\mathbf{h}_i[n]$

where $\mathbf{Q}_i[n]$ is a $\chi \times \chi$ unitary matrix, which accounts for the change of dominant eigenvectors without changing the subspace. Substituting (18) into (24), it can be easily verified that the MSE of the channel estimator is invariant to rotation of the dominant eigenvectors, which can also be seen in [26]. Therefore, it is the dominant subspace spanned by channel frequency responses that affects the performance of the subspace tracking based channel estimator.

3.3 Simulation Results for Space-Time Coded Systems

We use *typical-urban* (TU) delay profile with Doppler frequency f_d of 40 and 100 Hz in our simulation and assume channel parameters corresponding to different transmit and receive antenna pairs are independent but with the same statistical properties. The system has 2 transmit and 2 receive antennas. A total bandwidth of 1.25 MHz is divided into 256 subcarriers, 2 subcarriers on each end are used as guard tones, leaving the rest for transmitting data. The symbol duration is 204.8 μs and another 20.2 μs is added as cyclic prefix, resulting in a total block duration of 225 μs. A 16-state 2 b/s/Hz 4-PSK space-time code proposed in [33] is used in our simulation. Each data block of 500 bits is coded into two symbol blocks to be sent via the two transmit antennas at different subcarriers. 4 bits are used for trellis termination.

In each independent simulation, 2000 OFDM blocks of data are transmitted with 1 training block sent every M=10 blocks. The performance averaged over independent simulations is evaluated. For channel estimator with 7-tap subspace-tracking, the first 50 blocks use 9-tap DFT-based estimator. The estimated channel parameters are used for initial subspace acquisition so that initial training overhead can be saved at the expense of negligible performance loss for continuous data transmission. Then channel estimation is switched to the estimator with subspace tracking and the subspace is updated at each training block. The forgetting factor α is chosen to be 0.99.

Figures 9 and 10 show the average WER and *mean square error* (MSE) of the proposed estimator for system with Doppler frequency of 40 and 100 Hz, respectively. The performance of DFT-based robust estimator and that with ideal *channel state information* (CSI) are provided for comparison. It can be seen that for a WER of 1%, the performance of the proposed channel estimator has about 2 dB improvement over the DFT-based channel estimator for $f_d = 40$ Hz. This performance improvement is 1.8 dB for $f_d = 100$ Hz. The performance improvement can be clearly explained by the considerable MSE gap between two estimators at high SNR's given by Figures 9.b and 10.b, when MSE becomes dominated by leakage. Note that the MSE of proposed estimator is worse than that of the DFT-bassed estimator at low SNR's, which can be explained by the subspace tracking error caused by strong noise.

3.4 Channel Estimation for D-BLAST OFDM Systems Based on Subspace-Tracking

3.4.1 D-BLAST OFDM System

First, we briefly describe D-BLAST for MIMO-OFDM in this subsection.

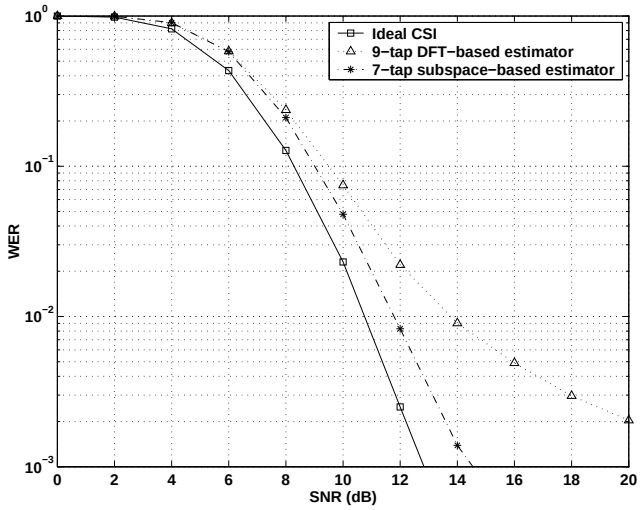
Equation (8) can also be written in matrix form as:

$$\mathbf{r}[n,k] = \mathbf{H}[n,k]\,\mathbf{s}[n,k] + \mathbf{w}[n,k], \tag{19}$$

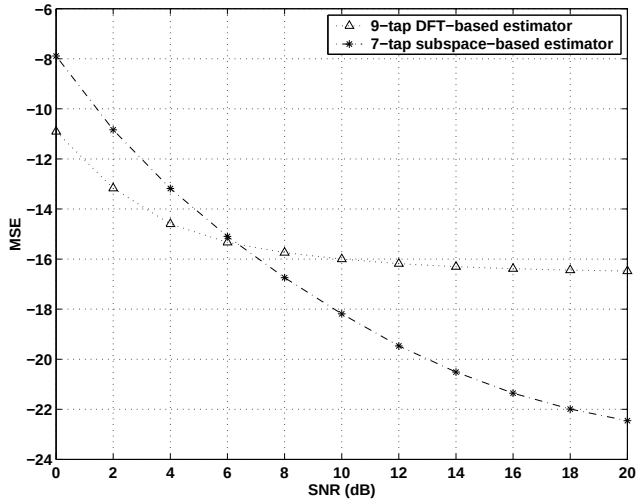
where

$$\mathbf{r}[n,k] = \begin{pmatrix} r_1[n,k] \\ \vdots \\ r_{N_r}[n,k] \end{pmatrix},$$

$$\mathbf{H}[n,k] = \begin{pmatrix} H_{11}[n,k] & H_{12}[n,k] & \cdots & H_{1N_t}[n,k] \\ H_{21}[n,k] & H_{22}[n,k] & \cdots & H_{2N_t}[n,k] \\ \vdots & \cdots & \ddots & \vdots \\ H_{N_r1}[n,k] & \cdots & \cdots & H_{N_rN_t}[n,k] \end{pmatrix},$$



(a)

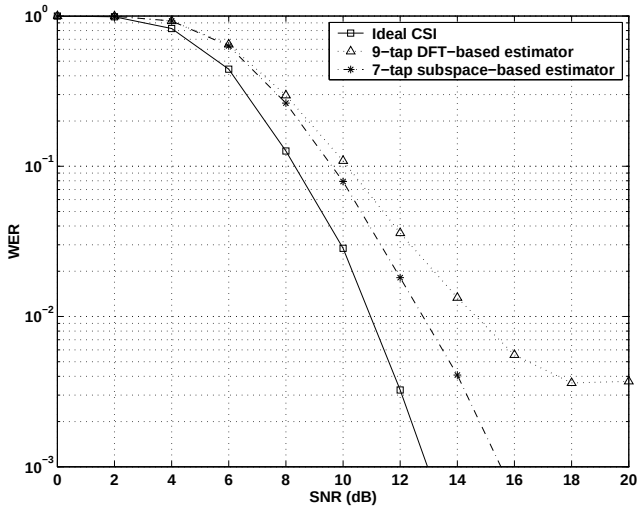


(b)

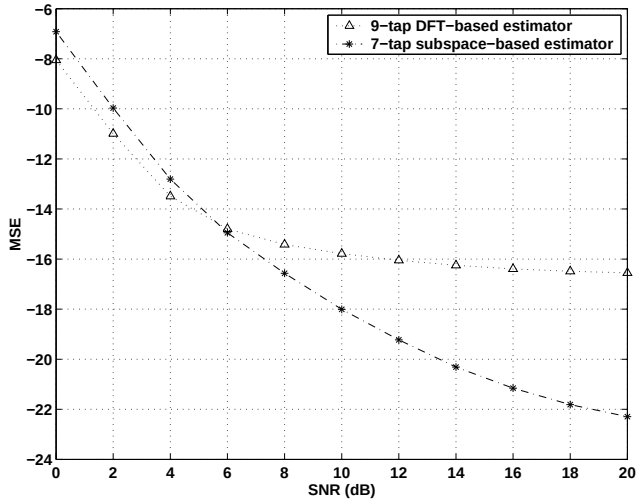
Figure 9: (a) WER and (b) MSE of MIMO-OFDM systems for channels with TU delay profile and f_d=40 Hz.

$$\mathbf{s}\left[n,k\right]=\left(\begin{array}{c} s_1\left[n,k\right] \\ \vdots \\ s_{N_t}\left[n,k\right] \end{array}\right), \text{ and } \mathbf{w}\left[n,k\right]=\left(\begin{array}{c} w_1\left[n,k\right] \\ \vdots \\ w_{N_r}\left[n,k\right] \end{array}\right).$$

D-BLAST is an effective MIMO technique [6] that has been originally developed for a single-carrier system with flat-fading channel. In this section, we will use this technique for a MIMO-OFDM system, which can be shown in Figure 11. From the figure, the set of all subcarriers in an OFDM block is divided into N_t subsets, each with $L = K/N_t$ subcarriers. Each layer, composed of N_t such subsets associated



(a)



(b)

Figure 10: (a) WER and (b) MSE of MIMO-OFDM systems for channels with TU delay profile and f_d=100 Hz.

with different transmit antennas, is encoded and decoded independently. Note that each layer still has K subcarriers, but different subcarriers may be associated with different transmit antennas. Layers starting at block n are denoted as $L_p[n]$, $p = 1, 2, \cdots, N_t$. With some abuse of notations, k is both the subcarrier index and the symbol index for each layer. Given the structure of D-BLAST OFDM, the received signal at each receive antenna is the superposition of the desired signal, the signals already detected in the previous layers, and those undetected.

The signal detection of D-BLAST MIMO-OFDM is also very similar to the original

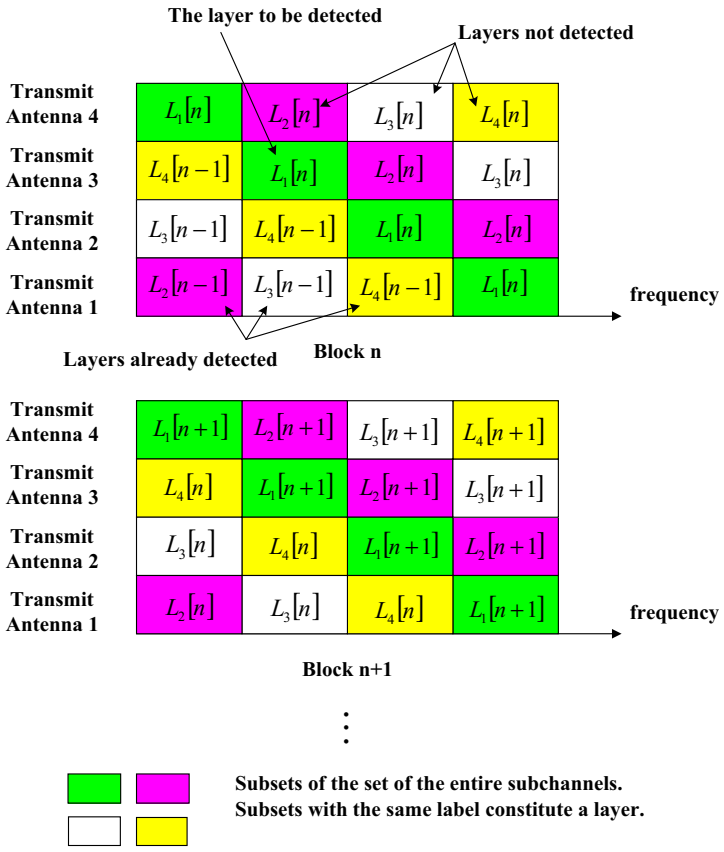


Figure 11: D-BLAST MIMO-OFDM structure.

D-BLAST. Assume layer $L_p[n]$, $p = 1,\ 2,\ \cdots,\ N_t$ is to be detected. From equation (8) we have:

$$\begin{aligned} r_j\left[g_p\left(n,k\right),k\right] \ = \ & H_{f_p(k),j}\left[g_p\left(n,k\right),k\right] s_{f_p(k)}\left[g_p\left(n,k\right),k\right] \\ & +\sum_{i=1}^{f_p(k)-1} H_{ij}\left[g_p\left(n,k\right),k\right] s_i\left[g_p\left(n,k\right),k\right] \\ & +\sum_{i=f_p(k)+1}^{N_t} H_{ij}\left[g_p\left(n,k\right),k\right] s_i\left[g_p\left(n,k\right),k\right] \\ & +w_j\left[g_p\left(n,k\right),k\right], \end{aligned} \tag{20}$$

for $j = 1,\ 2,\ \cdots,\ N_r$ and $k = 0,\ \ldots,\ K-1$, where $f_p(k)$ and $g_p(n,k)$ are associations of the k-th symbol of layer $L_p[n]$ with transmit antenna and OFDM block, respectively, *i.e.*, the k-th symbol of layer $L_p[n]$ is sent from the $f_p(k)$-th transmit antenna via the k-th subcarrier of the $g_p(n,k)$-th OFDM block. Note that in general a layer spans two consecutive OFDM blocks, thus $g_p(n,k)$ is either n or n+1. Equation (20) can be written in matrix notation as:

$$\begin{aligned}\mathbf{r}\left[g_p(n,k),k\right] = {} & \mathbf{H}_{f_p(k)}\left[g_p(n,k),k\right] s_{f_p(k)}\left[g_p(n,k),k\right] \\ & + \sum_{i=1}^{f_p(k)-1} \mathbf{H}_i\left[g_p(n,k),k\right] s_i\left[g_p(n,k),k\right] \\ & + \sum_{i=f_p(k)+1}^{N_t} \mathbf{H}_i\left[g_p(n,k),k\right] s_i\left[g_p(n,k),k\right] \\ & + \mathbf{w}\left[g_p(n,k),k\right],\end{aligned}$$

where $\mathbf{H}_i[n,k]$ is the i-th column of $\mathbf{H}[n,k]$. Signals from antennas 1 to $f_p(k)-1$ have been detected and those from antennas $f_p(k)+1$ to N_t are yet to be detected.

First, interference cancellation is carried out by subtracting detected signals from the received signal:

$$\begin{aligned}\tilde{\mathbf{r}}_p\left[g_p(n,k),k\right] = {} & \mathbf{r}\left[g_p(n,k),k\right] \\ & - \sum_{i=1}^{f_p(k)-1} \mathbf{H}_i\left[g_p(n,k),k\right] \hat{s}_i\left[g_p(n,k),k\right],\end{aligned}$$

where $\hat{s}_i[n,k]$'s are detected symbols. Then interference from undetected signals is suppressed by linear combination that yields the maximum *signal-to-interference-plus-noise ratio* (SINR). Let

$$\tilde{\mathbf{H}}_p[n,k] \triangleq \left(\mathbf{H}_{f_p(k)+1}[n,k],\ \mathbf{H}_{f_p(k)+2}[n,k],\ \cdots,\ \mathbf{H}_{N_t}[n,k]\right),$$

then from [34], we have the following weighting vector:

$$\mathbf{v}_p[n,k] = \left(\tilde{\mathbf{H}}_p[n,k]\,\tilde{\mathbf{H}}_p^H[n,k] + \rho\mathbf{I}\right)^{-1} \mathbf{H}_{f_p(k)}[n,k].$$

Thus, if we assume Gaussian distribution for the residual interference plus noise, the maximum likelihood decoding of layer $L_p[n]$ is to find $\left\{\hat{s}_{f_p(k)}\left[g_p\left(n,k\right),k\right]\right\}$ that minimizes

$$\begin{aligned}
&M\left(\left\{\hat{s}_{f_p(k)}\left[g_p\left(n,k\right),k\right];\ k=0,\ 1,\ \cdots,\ K-1\right\}\right)\\
=&\sum_{k=0}^{K-1}\frac{1}{\mathbf{v}_p^H[m,k]\left(\tilde{\mathbf{H}}_p[m,k]\,\tilde{\mathbf{H}}_p^H[m,k]+\rho\mathbf{I}\right)\mathbf{v}_p[m,k]}\\
&\cdot\left|\mathbf{v}_p^H[m,k]\left(\tilde{\mathbf{x}}_p[m,k]-\mathbf{H}_{f_p(k)}[m,k]\,\hat{s}_{f_p(k)}[m,k]\right)\right|^2\Big|_{m=g_p(n,k)}.
\end{aligned}$$

which can be solved by standard Viterbi algorithm when convolutional codes are used. From the above discussions, channel information is crucial for the signal detection of D-BLAST MIMO-OFDM.

3.4.2 Layer-Wise Least-Square Channel Estimation

Due to layer-wise detection in D-BLAST, usually only partial knowledge of the symbols transmitted from all transmit antennas at one OFDM block is available after decoding of each layer. To exploit the characteristics of D-BLAST structure, channel estimation is carried out each time a layer is detected.

After detection of layer $L_p[n]$, we estimate the channel responses at the n-th block according to Equation (8). Since only part of all the subcarriers of the current OFDM block have signals from all transmit antennas detected, we replace the received signals at subcarriers not fully detected with those of the previous OFDM block to form a complete received signal vector, due to the fact that $H_i[n,k]\approx H_i[n-1,k]$. Define

$$z^{(p)}[n,k]=\begin{cases}r[n-1,k], & k\in\Sigma^{(p)},\\ r[n,k], & \text{else.}\end{cases}$$

and

$$\hat{d}_i^{(p)}[n,k]\triangleq\begin{cases}\hat{s}_i[n-1,k], & k\in\Sigma^{(p)},\\ \hat{s}_i[n,k], & \text{else,}\end{cases}$$

where $\Sigma^{(p)} = \{k;\ g_p(n,k) = n \text{ and } f_p(k) < N_t\}$ is the set of subcarriers with signals not fully detected.

Let

$$\begin{aligned}
\mathbf{z}^{(p)}[n] &= \left(z^{(p)}[n,0],\ \cdots,\ z^{(p)}[n,K-1]\right)^T, \\
\hat{\mathbf{D}}_i^{(p)}[n] &= diag\left\{\hat{d}_i^{(p)}[n,0],\ \cdots,\ \hat{d}_i^{(p)}[n,K-1]\right\}, \\
\hat{\mathbf{h}}^{(p)}[n] &\triangleq \left(\hat{\mathbf{h}}_1^{(p)^T}[n],\ \hat{\mathbf{h}}_2^{(p)^T}[n],\ \cdots,\ \hat{\mathbf{h}}_{N_t}^{(p)^T}[n]\right)^T,
\end{aligned} \tag{21}$$

and

$$\hat{\mathbf{h}}_i^{(p)}[n] \triangleq \left(\hat{h}_i^{(p)}[n,0],\ \cdots,\ \hat{h}_i^{(p)}[n,\chi-1]\right)^T.$$

The least-square channel estimation is to minimize the following cost function [20]:

$$\begin{aligned}
&C\left(\left\{\hat{\mathbf{h}}_i^{(p)};\ i=1,\ 2,\ \cdots,\ N_t-1\right\}\right) \\
&= \left\|\mathbf{z}^{(p)}[n] - \sum_{i=1}^{N_t} \hat{\mathbf{D}}_i^{(p)}[n]\,\mathbf{U}\hat{\mathbf{h}}_i^{(p)}[n]\right\|^2.
\end{aligned}$$

Then

$$\hat{\mathbf{h}}^{(p)}[n] = \left(\mathbf{T}^{(p)^H}[n]\,\mathbf{T}^{(p)}[n]\right)^{-1}\mathbf{T}^{(p)^H}[n]\,\mathbf{z}^{(p)}[n] \tag{22}$$

where

$$\mathbf{T}^{(p)}[n] = \left(\hat{\mathbf{D}}_1^{(p)}[n]\,\mathbf{U},\ \hat{\mathbf{D}}_2^{(p)}[n]\,\mathbf{U},\ \cdots,\ \hat{\mathbf{D}}_{N_t}^{(p)}[n]\,\mathbf{U}\right). \tag{23}$$

The above estimate is further refined by applying a robust estimator for OFDM systems in [26], which makes full use of the correlation of channel parameters at different OFDM blocks.

3.4.3 Performance Analysis

Here, we briefly analyze the performance of the above channel estimator for D-BLAST OFDM. Let

$$\Delta\mathbf{h} \triangleq \mathbf{h}[n] - \mathbf{h}[n-1],$$

$$\hat{s}_i^{(p)}[n,k] \triangleq \begin{cases} \hat{b}_i[n-1,k], & k \in \Sigma^{(p)}, \\ 0, & \text{else}, \end{cases}$$

$$\hat{\mathbf{S}}_i^{(p)}[n] = diag\left\{\hat{s}_i^{(p)}[n,0], \cdots, \hat{s}_i^{(p)}[n,K-1]\right\},$$

and

$$\mathbf{G}^{(p)}[n] \triangleq \left(\mathbf{T}^{(p)^H}[n]\,\mathbf{T}^{(p)}[n]\right)^{-1}\mathbf{T}^{(p)^H}[n]\,\hat{\mathbf{S}}_i^{(p)}[n].$$

The mean-square error of the channel estimator is

$$\begin{aligned} MSE^{(p)}[n] &\triangleq \frac{1}{N_t\chi}E\left\{\left\|\hat{\mathbf{h}}^{(p)}[n]-\mathbf{h}[n]\right\|^2\right\} \\ &= \frac{1}{N_t\chi}\mathbf{Tr}\left\{\rho\left(\mathbf{T}^{(p)^H}[n]\,\mathbf{T}^{(p)}[n]\right)^{-1}\right. \\ &\quad \left.+\mathbf{G}^{(p)}[n]\,E\left(\Delta\mathbf{h}[n]\,\Delta\mathbf{h}^H[n]\right)\mathbf{G}^{(p)^H}[n]\right\}, \end{aligned} \tag{24}$$

where $E\{\cdot\}$ denotes expected value of a random variable. Clearly the first term in the above equation results from noise and the second term is due to channel variation. Following the argument in Section 3.2, we can further improve channel estimation accuracy by using the optimum transform matrix derived from tracking the dominant subspace spanned by channel frequency responses, thus reducing the error floor caused by leakage.

3.4.4 Simulation Results for D-BLAST OFDM

In this section we evaluate the performance of different decision-directed channel estimation algorithms for D-BLAST OFDM by computer simulation. TU channel with Doppler frequency $f_d = 40$ and 200 Hz is used in our simulation. Performance of the proposed 7-tap layer-wise subspace-tracking estimator is simulated. As a comparison, performances of systems with ideal channel parameters, 7-tap layer-wise estimator with optimum transform, as defined in Equation (17), and 10-tap layer-wise DFT-based estimator with *significant tap selection* (STS) [20] are evaluated. The performance of the traditional 10-tap block-wise DFT-based channel estimator

is also given, where channel estimation is carried out once per OFDM block and the estimated channel parameters are used for the detection of the next OFDM block.

Four transmit antennas and four receive antennas are employed to form four D-BLAST layers. Channel parameters corresponding to different transmit and receive antenna pairs are assumed to be independent but have the same statistics. The system bandwidth of 1.25 MHz is divided into 256 subchannels. 2 subchannels on each side are used as guard tones, and the rest of the subchannels are used for data transmission. The symbol duration is 204.8 μs and another 20.2 μs is added as cyclic prefix, resulting in a total block duration of 225 μs. A 16-state binary-to-4-ary convolutional codes of rate 1/2 with the octal generators being (26,37) [35] is used to encode the information bits in each layer. Four tail bits are used for trellis termination, leaving 248 information bits per layer. The encoder output is interleaved before sending to a transmit antenna at a particular subcarrier.

In each independent simulation, 2000 OFDM blocks of data are transmitted with 1 training block sent every 10 blocks. The performance averaged over independent simulations is evaluated. For channel estimator with subspace-tracking, the first 50 blocks use 10-tap DFT-based estimator with STS. The estimated channel parameters are used for initial subspace acquisition so that initial training overhead can be saved at the expense of negligible performance loss for continuous data transmission. Then channel estimation is switched to the estimator with subspace tracking and the subspace is updated at each training block. The forgetting factor α is chosen to be 0.995.

Figure 12 (a) and (b) compare the WER and *bit-error-rate* (BER) performance of different channel estimation algorithms when Doppler frequency is 40 Hz. Of all estimators, the block-wise DFT-based channel estimator has the worst performance since it uses channel state information at the previous OFDM block for detection and thus is most susceptible to channel variation. The block-wise DFT-based estimator

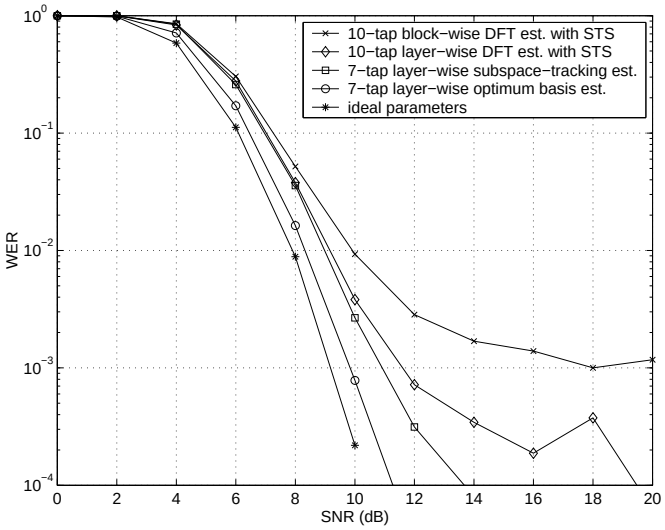
requires about 1 dB more SNR than the layer-wise DFT-based estimator for a WER of 10^{-2} and its WER curve levels off quickly at high SNR's since its performance is bounded by channel variation. Among layer-wise estimators, the subspace-tracking estimator requires 0.7 dB less SNR than the DFT-based estimator for 10^{-3} WER. Figure 12 (c) shows how MSE evolves as the layer-wise channel estimation progresses. From the figure we can see that for all layer-wise channel estimation methods, the most significant MSE improvement is seen after detection of the first layer of the current OFDM block, which is 0.7 dB at SNR=16 dB, compared with about 0.16 dB per layer improvement for layers detected later with the proposed subspace-tracking channel estimator.

For f_d=200 Hz, from Figure 13 we see that the performance difference between the block-wise channel estimator and layer-wise estimators is even bigger now that the system performance is dominated by fast variation of channel parameters. The SNR gain for using layer-wise subspace-tracking estimator is about 0.8 dB for 10^{-2} WER compared with layer-wise DFT-based estimator. It is clear that as the channel variation rate increases, the MSE performance improvement with layer-wise channel estimation becomes more significant, with the successive MSE improvements being 3.4 dB, 1.2 dB, 1.2 dB, and 1 dB at SNR=16 dB, as observed in Figure 13 (c). For both $f_d = 40$ and 200 Hz, the subspace-tracking estimator can effectively reduce the error floor thus provide better performance than that of the DFT-based estimator.

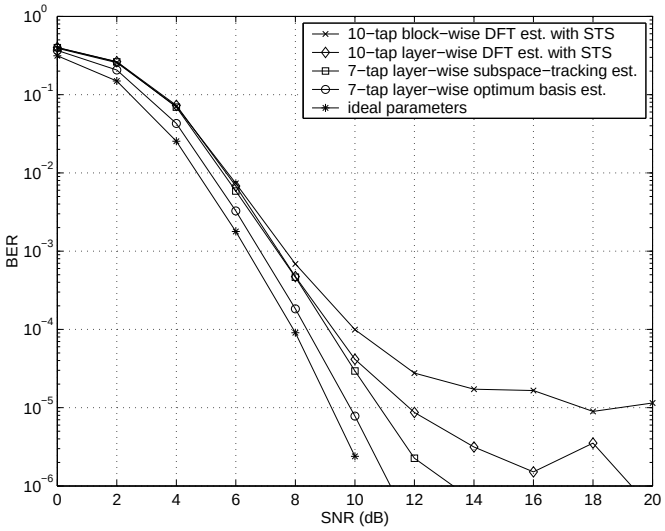
3.5 Summary

In this chapter, we propose a subspace-based channel estimation algorithm for MIMO-OFDM systems. By tracking the slow varying subspace spanned by the the frequency responses of different multipaths, we obtain the optimum basis for LS channel estimation. The proposed algorithm considerably reduces the leakage of low-rank channel estimation, which is dominant at high SNR's and the corresponding performance

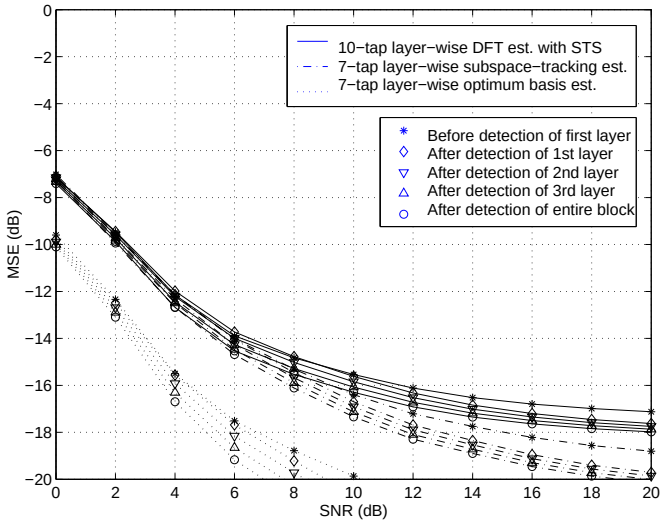
improvement is shown by simulation. Next, we apply the D-BLAST structure to MIMO-OFDM systems and develop a channel estimator that updates the estimated channel parameters in a layer-wise fashion. Since we update channel estimation using detected signals to improve detection of the rest of the signals in the current OFDM block, the system is more robust to fast fading channels, when compared with the traditional block-wise estimator. To further reduce the channel estimation error, we use the training blocks not only for channel estimation, but also for tracking of the dominant subspace spanned by the channel frequency response to reduce the number of parameters to be estimated during data transmission mode. Thus, additional performance improvement is obtained by using subspace tracking for the layer-wise estimator, which is about 0.8 dB for 10^{-2} WER with $f_d = 200$ Hz.



(a)

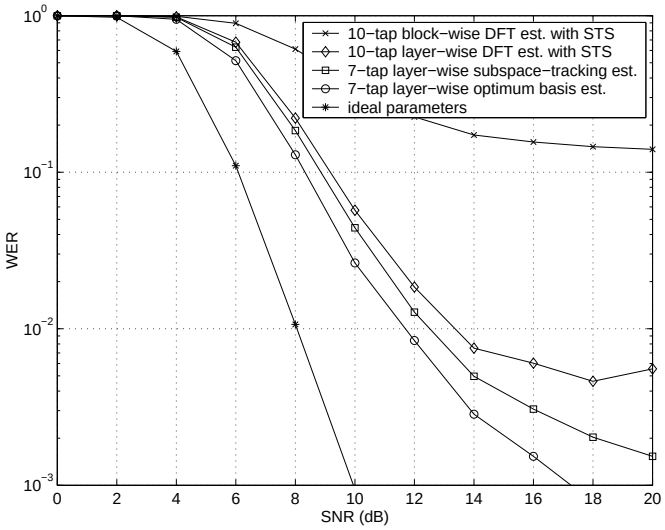


(b)

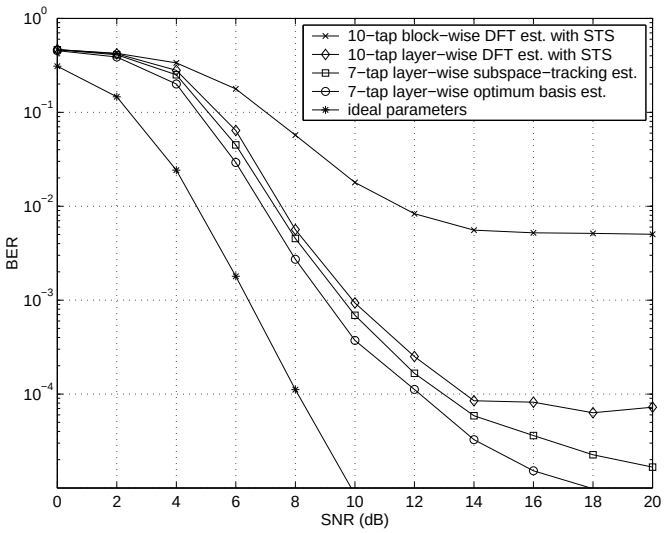


(c)

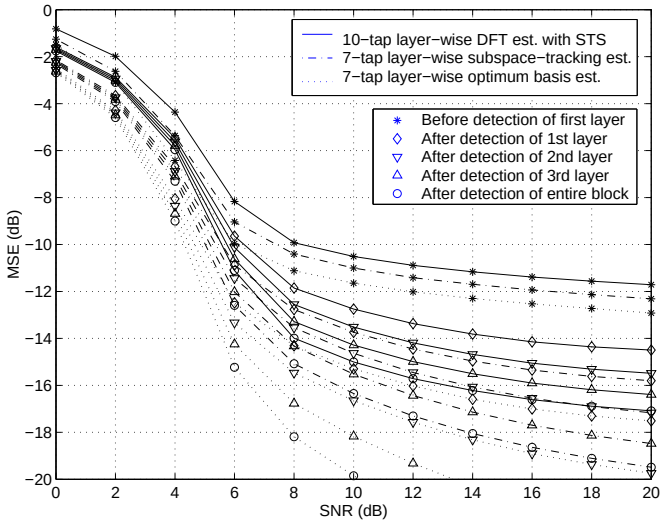
Figure 12: (a) WER and (b) BER (c) MSE of D-BLAST systems for channels with TU delay profile and f_d=40 Hz.



(a)



(b)



(c)

Figure 13: (a) WER and (b) BER (c) MSE of D-BLAST systems for channels with TU delay profile and f_d=200 Hz.

CHAPTER IV

PARALLEL DETECTION WITH PREDICTIVE SOFT INTERFERENCE CANCELLATION

Given the advantages and disadvantages of STC and V-BLAST as discussed in Chapter 2, group-wise space-time coding is proposed as a tradeoff between complexity and performance. It can be treated as generalized layered structure where transmit antennas are divided into small groups, and independent space-time codes, called *component codes*, are used to transmit information for each group [17], [36]. In [17], the received signal vector is projected into the orthogonal complement of the subspace spanned by the interfering signals, thus completely suppressing the interference. In [36], the received signal vector is spatially pre-whitened followed by *minimum-euclidean-distance* (MED) decoding, so that the component of the desired signal vector in the subspace spanned by the interfering signals is also taken into account by incorporating the reliability information. The detection algorithm in [36] is similar in assuming that symbols from the interfering antennas are independent with zero-mean. We propose a parallel detection algorithm, and it is shown that by estimating the a posteriori probabilities of the states of the component code trellis and exploiting the structure of the space-time codes in [3], system performance can be considerably improved.

A wireless communication system with four antennas at the transmitter and four antennas at the receiver is shown in Figure 14. Extension to systems with different number of antennas is straightforward.

When only detection is concerned, we may drop the subcarrier index k in Equation

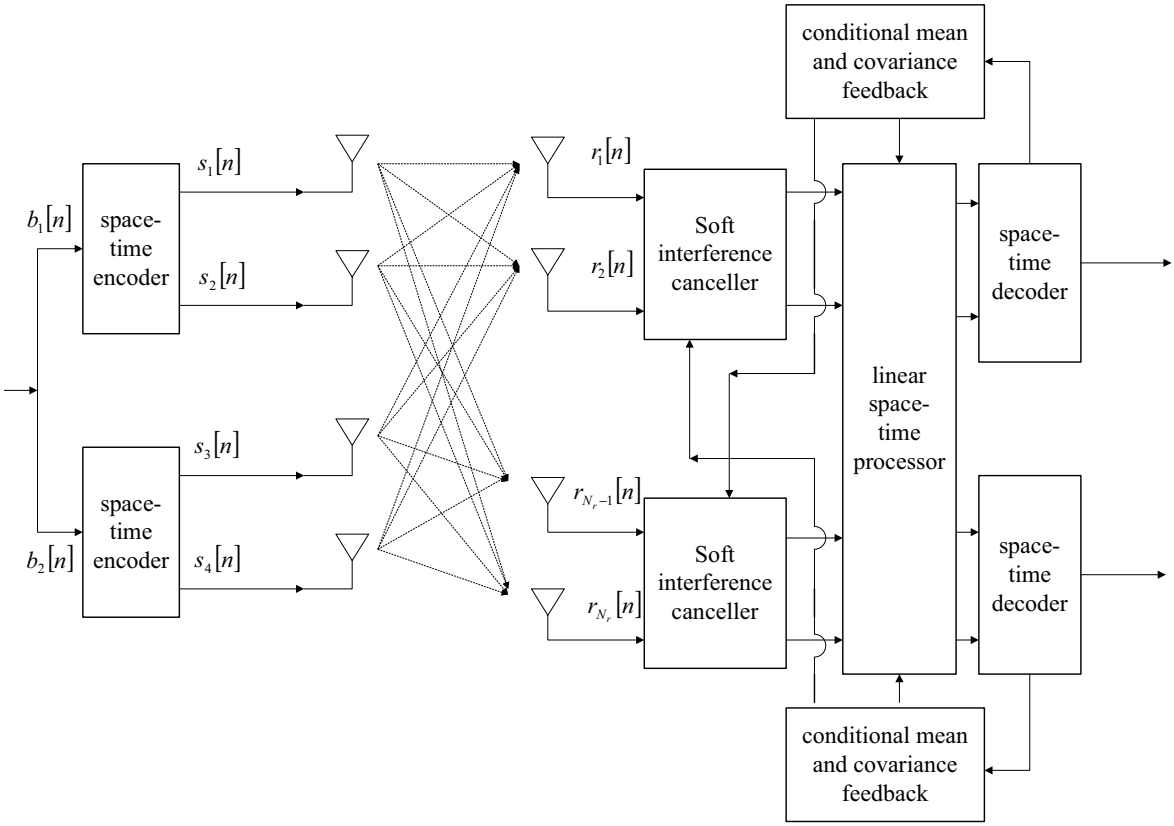


Figure 14: A MIMO system with two component space-time codes.

(8) without loss of generality and have it expressed in matrix notation as

$$\mathbf{r}\left[n\right] = \mathbf{H}_1\left[n\right]\mathbf{s}_1\left[n\right] + \mathbf{H}_2\left[n\right]\mathbf{s}_2\left[n\right] + \mathbf{w}\left[n\right], \tag{25}$$

where

$$\mathbf{r}\left[n\right] \triangleq \begin{pmatrix} r_1\left[n\right] \\ \vdots \\ r_{N_r}\left[n\right] \end{pmatrix}, \ \mathbf{s}_i\left[n\right] \triangleq \begin{pmatrix} s_{2i-1}\left[n\right] \\ s_{2i}\left[n\right] \end{pmatrix}, \ \mathbf{w}\left[n\right] \triangleq \begin{pmatrix} w_1\left[n\right] \\ \vdots \\ w_{N_r}\left[n\right] \end{pmatrix},$$

and

$$\mathbf{H}_i\left[n\right] \triangleq \begin{pmatrix} H_{1,2i-1}\left[n\right] & H_{1,2i}\left[n\right] \\ \vdots & \vdots \\ H_{N_r,2i-1}\left[n\right] & H_{N_r,2i}\left[n\right] \end{pmatrix}.$$

4.1 Signal Detection Using Predictive Soft Interference Cancellation

To simplify detection complexity, when one component code is to be decoded, the transmitted signal from the other one is treated as interference instead of joint detection. The difference of the proposed detection algorithm from that in [36] is that we no longer take $\mathbf{s}_2[n]$'s as spatially white with zero-mean. Instead, we use the partially decoded information from the Viterbi decoder to calculate the a posteriori probabilities of different trellis states of component codes given previously received signals. The motivation is that by exploiting the particular trellis structure of the space-time codes in [3] and exchanging information between the two component space-time decoder, we may be able to reduce the uncertainty in self interference, and thus improve system performance. As shown in Figure 15, all branches from the same state have one output symbol in common. As we proceed along the trellis during Viterbi decoding, we can calculate the probability of each state given all previous observed signals. With the probability distribution of the states at time index n, we calculate probabilities of output symbols at time index $n+1$ by assuming all branches from one state are equally likely. When SNR is high, one state is much more likely than other states, and we can cancel much of the interference to the other component code from the output symbol shared by all branches from that state by subtracting the conditional mean of the interference. With soft interference cancellation rather than hard interference cancellation, we reduce the detrimental effect of error propagation due to incorrect detection of the interfering signal.

First, we carry out soft interference cancellation based on the previously observed signals and the structure of the space-time code trellis.

$$\mathbf{r}_1[n] = \mathbf{r}[n] - \mathbf{H}_2[n]\,\bar{\mathbf{s}}_2[n], \tag{26}$$

where $\bar{\mathbf{s}}_2[n]$ is the conditional mean of the interfering signal given information from the

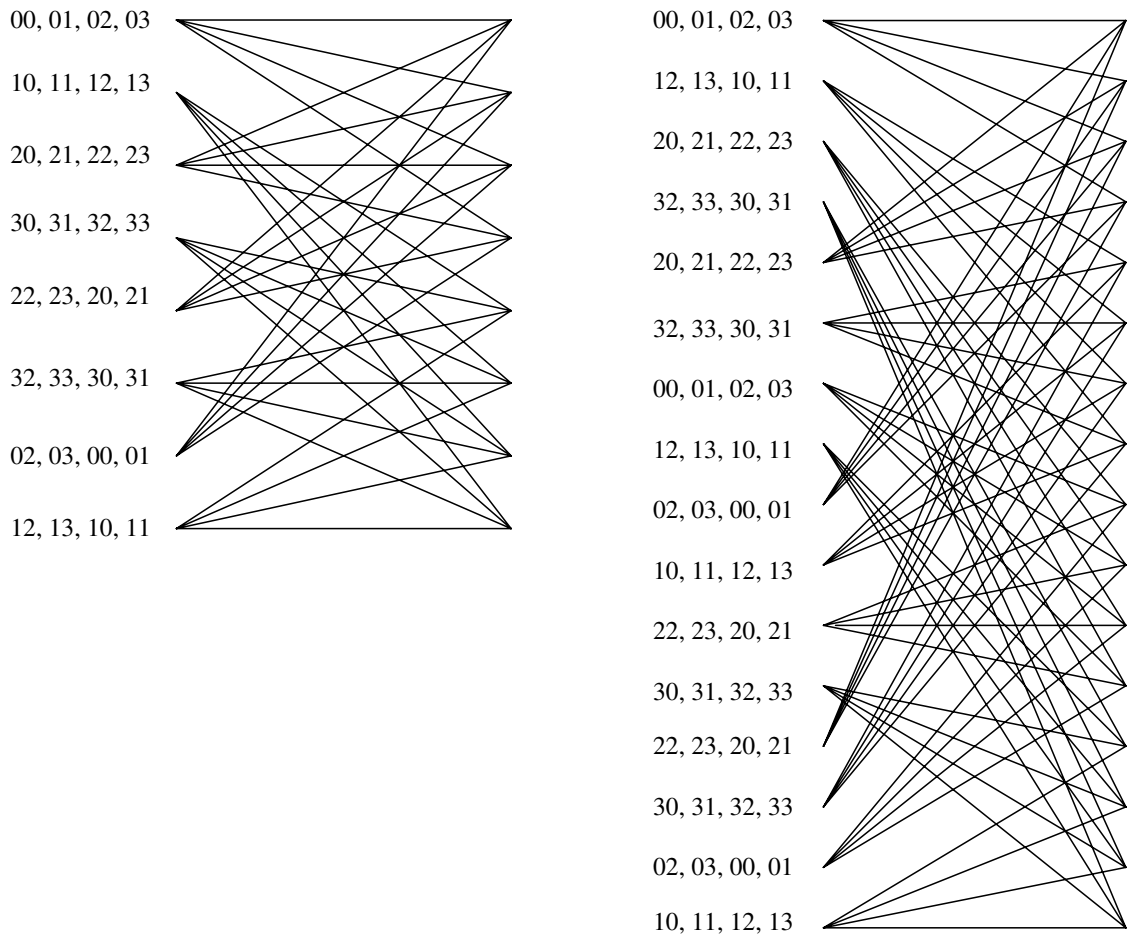


Figure 15: 2-space-time codes with 8 and 16 states, 4PSK.

decoder for the second component space-time code after $\mathbf{r}[0], \mathbf{r}[1], \cdots, \mathbf{r}[n-1]$.

Thus, to detect $\mathbf{s}_1[n]$, the received signal can be expressed as

$$\mathbf{r}_1[n] = \mathbf{H}_1[n]\mathbf{s}_1[n] + \mathbf{v}_1[n], \tag{27}$$

where $\mathbf{v}_1[n] = \mathbf{H}_2[n](\mathbf{s}_2[n] - \bar{\mathbf{s}}_2[n]) + \mathbf{w}[n]$.

Similar to [36], we apply a spatial pre-whitener. After the soft interference cancellation, let the covariance matrix of the residual interfering symbols be $\mathbf{C}_2[n]$.

Define

$$\mathbf{R}_1[n] \triangleq E\left\{\mathbf{v}_1[n]\mathbf{v}_1^H[n]\right\} = \mathbf{H}_2[n]\mathbf{C}_2[n]\mathbf{H}_2^H[n] + \rho\mathbf{I}.$$

Since $\mathbf{R}_1[n]$ is positive definite, by Cholesky factorization [27], there exists a lower triangular matrix $\mathbf{L}$ with non-negative diagonal entries such that

$$\mathbf{L}_1[n]\mathbf{L}_1^H[n] = \mathbf{R}_1[n].$$

Use $\mathbf{L}_1^{-1}[n]$ to whiten the received signal. From (27), we have

$$\hat{\mathbf{r}}_1[n] = \hat{\mathbf{H}}_1[n]\mathbf{s}_1[n] + \hat{\mathbf{v}}_1[n],$$

where

$$\hat{\mathbf{r}}_1[n] = \mathbf{L}_1^{-1}[n]\mathbf{r}_1[n],$$

$$\hat{\mathbf{H}}_1[n] = \mathbf{L}_1^{-1}[n]\mathbf{H}_1[n],$$

and

$$\hat{\mathbf{v}}_1[n] = \mathbf{L}_1^{-1}[n]\mathbf{v}_1[n].$$

Note that $\hat{\mathbf{H}}_1[n]$ is the virtual channel matrix after the spatial pre-whitener. Assuming that all codewords are equiprobable, the decoding is carried out by finding a valid codeword $\{\mathbf{s}_1[n]\}$ that minimize the decision metric

$$\sum_n \left|\hat{\mathbf{r}}_1[n] - \hat{\mathbf{H}}_1[n]\mathbf{s}_1[n]\right|^2. \tag{28}$$

Viterbi algorithm can be used for decoding.

Now we will show how to calculate $\bar{\mathbf{s}}_1[n]$ and $\mathbf{C}_1[n]$, information to be fed to the 2nd component space-time decoder. $\bar{\mathbf{s}}_2[n]$ and $\mathbf{C}_2[n]$ can be derived in the same way. It should be noted that $\bar{\mathbf{s}}_1[n]$ is not exactly the conditional mean given $\mathbf{r}[0], \mathbf{r}[1], \cdots, \mathbf{r}[n-1]$ since the metric used for calculating the probability distribution of $\mathbf{s}_1[n]$ uses the output of the spatial whitener where independent Gaussian distribution of the interference is assumed. Denote the state of the first component space-time encoder at time n as $S[n]$, and the output corresponding to the transition from state m' to state m as $\mathbf{s}_{m'm}$. With some abuse of notations, we use $\Pr\{\cdot\}$ to denote the probability functions calculated from the metric in (28). Define

$$\mu_m[n] \triangleq \Pr\{S[n] = m \; ; \; \hat{\mathbf{r}}_1[0] \; \hat{\mathbf{r}}_1[1] \cdots \hat{\mathbf{r}}_1[n]\},$$

and

$$\gamma_{m'm}[n] \triangleq \Pr\{S[n] = m; \; \hat{\mathbf{r}}_1[n] \; | \; S[n-1] = m'\}.$$

By the BCJR algorithm [37], we can calculate $\mu_m[n]$ recursively as

$$\mu_m[n] = \sum_{m'} \mu_{m'}[n-1] \cdot \gamma_{m'm}[n].$$

The initial condition is $\mu_0[0] = 1$ and $\mu_m[0] = 0$ for $m \neq 0$ since the encoder begins with the zero state.

With independent Gaussian distribution assumption for the interference,

$$\gamma_{m'm}[n] \sim e^{-\left\|\hat{\mathbf{r}}_1[n] - \hat{\mathbf{H}}_1[n]\mathbf{s}_{m'm}\right\|^2}.$$

Then the recursion is simply

$$\mu_m[n] = \sum_{m'} \mu_{m'}[n-1] \cdot e^{-\left\|\hat{\mathbf{r}}_1[n] - \hat{\mathbf{H}}_1[n]\mathbf{s}_{m'm}\right\|^2}.$$

since the constant factor is eliminated when calculating the conditional probability

$$\begin{aligned} & \Pr\{S[n] = m \mid \hat{\mathbf{r}}_1[0] \; \hat{\mathbf{r}}_1[1] \cdots \hat{\mathbf{r}}_1[n]\} \\ = \; & \frac{\Pr\{S[n] = m \; ; \; \hat{\mathbf{r}}_1[0] \; \hat{\mathbf{r}}_1[1] \cdots \hat{\mathbf{r}}_1[n]\}}{\sum_k \Pr\{S[n] = k \; ; \; \hat{\mathbf{r}}_1[0] \; \hat{\mathbf{r}}_1[1] \cdots \hat{\mathbf{r}}_1[n]\}} \\ = \; & \frac{\mu_k[n]}{\sum_k \mu_k[n]}. \end{aligned} \tag{29}$$

Assume transitions from the same state are equiprobable. With the a posteriori probability associated with each state at time $n-1$, we can compute the conditional mean and covariance of the signal at time n as

$$\bar{\mathbf{s}}_1[n] = \sum_{m',m} \frac{\delta(m',m)\,\mu_m[n-1]}{N_B \sum_k \mu_k[n-1]} \mathbf{s}_{m'm},$$

and

$$\begin{aligned} \mathbf{C}_1[n] &= E\left\{ (\mathbf{s}_2[n] - \bar{\mathbf{s}}_1[n]) (\mathbf{s}_2[n] - \bar{\mathbf{s}}_1[n])^H \right\}. \\ &= \sum_{m',m} \frac{\delta(m',m)\,\mu_m[n-1]}{N_B \sum_k \mu_k[n-1]} \\ &\quad \cdot (\mathbf{s}_{m'm} - \bar{\mathbf{s}}_1[n]) (\mathbf{s}_{m'm} - \bar{\mathbf{s}}_1[n])^H, \end{aligned}$$

where N_B is the number of branches starting from each state and

$$\delta(m',m) = \begin{cases} 1, & \text{there is a transition from state } m' \text{ to } m, \\ 0, & \text{else.} \end{cases}$$

As an example, the decoding trellis diagram for a 8-state 2-space-time code in [3] is shown in Figure 16.

The proposed detection algorithm is summarized in Table 3. Note that $\mu_m[n]$'s are normalized after each recursion to prevent the denominator in (29) becoming too small. The normalization does not change the conditional probabilities of different states since they are invariant to scaling of $\mu_m[n]$'s.

In other interference cancellation based algorithms such as those for BLAST or STBC in [38], [39], [40], [41] and [42], the reconstruction of interference in other schemes requires the decoding of the entire code word. Therefore, simultaneous decoding of all component codes is impossible. As a result, those schemes can only be implemented as successive interference cancellation or iterative detection. In contrast, the proposed algorithm exploits the partial information during Viterbi decoding and the structure of specific code trellises to predict interference statistics at the next

Table 3: Parallel detection of group-wise space-time coding using predictive soft interference cancellation

Initialization

$\mu_0[0] = 1$ and $\mu_m[0] = 0$ for $m \neq 0$

I Interference prediction

Computation of soft interference

$$\bar{\mathbf{s}}_1[n] = \sum_{m',m} \frac{\delta(m',m)\,\mu_m[n-1]}{N_B} \mathbf{s}_{m'm}$$

Computation of covariance matrix of residual interference

$$\mathbf{C}_1[n] = \sum_{m',m} \frac{\delta(m',m)\,\mu_m[n-1]}{N_B} (\mathbf{s}_{m'm} - \bar{\mathbf{s}}_1[n])(\mathbf{s}_{m'm} - \bar{\mathbf{s}}_1[n])^H$$

II Interference cancellation and pre-whitening at time n

Soft interference cancellation

$\mathbf{r}_1[n] = \mathbf{r}[n] - \mathbf{H}_2[n]\,\bar{\mathbf{s}}_2[n]$

Spatial pre-whitening

$\mathbf{R}_1[n] = \mathbf{H}_2[n]\,\mathbf{C}_2[n]\,\mathbf{H}_2^H[n] + \rho\mathbf{I}$

$\mathbf{L}_1[n]\,\mathbf{L}_1^H[n] = \mathbf{R}_1[n]$

$\hat{\mathbf{r}}_1[n] = \mathbf{L}_1^{-1}[n]\,\mathbf{r}_1[n]$

Computation of a posteriori probabilities of states at time n $(n \neq 0)$

$$\mu_m[n] = \sum_{m'} \mu_{m'}[n-1] \cdot e^{-\left\|\hat{\mathbf{r}}_1[n] - \hat{\mathbf{H}}_1[n]\mathbf{s}_{m'm}\right\|^2}$$

$$\mu_m[n] = \frac{\mu_m[n]}{\sum_k \mu_k[n]}$$

III MED decoding

$$\{\mathbf{s}_1[n]\} = \arg\min \sum_n \left\| \hat{\mathbf{r}}_1[n] - \hat{\mathbf{H}}_1[n]\,\mathbf{s}_1[n] \right\|^2$$

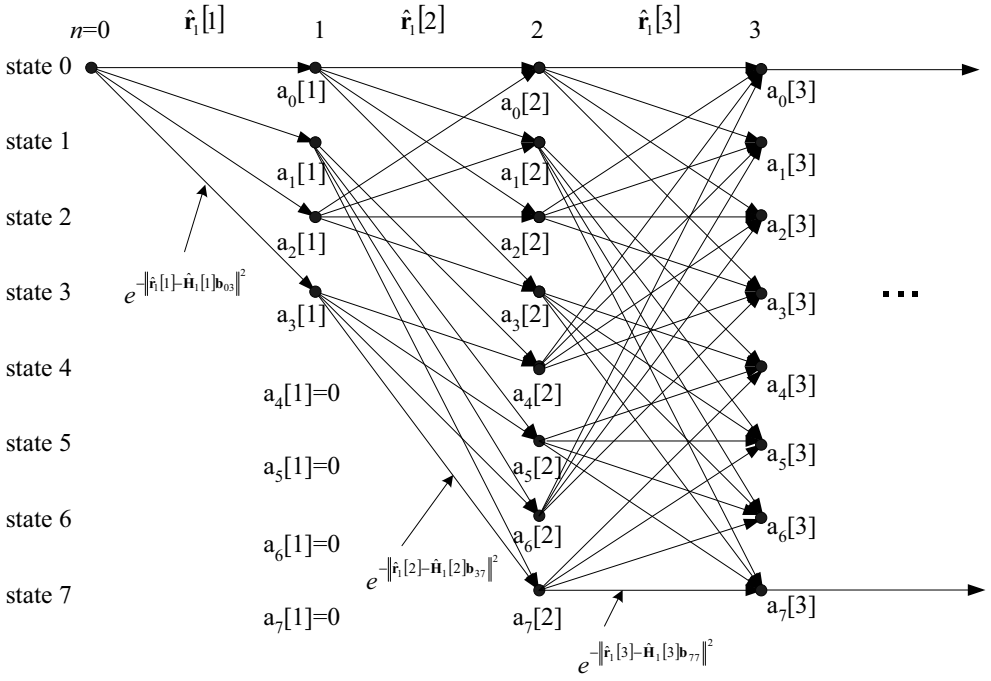


Figure 16: Trellis diagram for a 2-space-time code with 8 states, 4PSK.

time index. Therefore parallel interference cancellation can be implemented, which avoids the decoding delay common in successive interference cancellation schemes. In addition, the modularity of the proposed algorithm is especially suitable for modern VLSI implementation.

4.2 Simulation Results

We proceed to evaluate the performance of proposed algorithm in the previous section in a quasi-static flat-fading channel. The simulated system is equipped with four transmit and four receive antennas. Each two transmit antennas are grouped as one component space-time code. The two-space-time codes for 4PSK constellations shown in Figure 15 are used, delivering a total bandwidth efficiency of 4 b/s/Hz. The length of each code word is 256. The encoding of these trellis codes begins and ends with zero state. The WER's of the proposed algorithm are averaged over 10^5 independent simulations. The performance of the scheme that assumes interfering symbols are

spatially white with zero mean in [36] is also given for comparison.

WER curves for 2-space-time codes with 8 states are shown in Figure 17. For a 1% WER, the proposed algorithm for group-wise space-time coding reduces the required SNR the scheme in [36] by 1.7 dB. For 2-space-time codes with 16 states, the performance improvement is 1.5 dB, as shown in Figure 18.

The performance improvement is due to the reduction of self interference after interference cancellation. Or equivalently, the significant dimension of the subspace spanned by interference is reduced by interference cancellation, which implies more diversity can be gained for the desired component code. In the case of the 2-space-time codes used for simulation, with perfect determination of the present state (probability of the corresponding state being 1), which is the most favorable situation, 1 of the 2 symbols from the interfering component space-time code at the next time index can be completely determined according to the code trellis. Thus, its contribution to the interference can be completely eliminated by interference cancellation. Therefore, the upper bound for the dimension reduction of the subspace spanned by interference with proposed detection algorithm is one.

4.3 Summary

Group-wise space-time coding is a promising tradeoff between complexity and performance in achieving the attractive capacity of MIMO systems. We propose a novel parallel signal detection algorithm that exploits the structure of specific space-time code trellises to predict the interference for next time index during Viterbi decoding. In a quasi-static flat-fading channel with 4 transmit 4 receive antennas, the proposed algorithm improves the WER performance by 1.7 dB and 1.5 dB for 2-space-time codes with 8 and 16 states, respectively, for a WER of 1%. The performance improvement is due to the reduction of self interference after interference cancellation. Or equivalently, the significant dimension of the subspace spanned by interference is

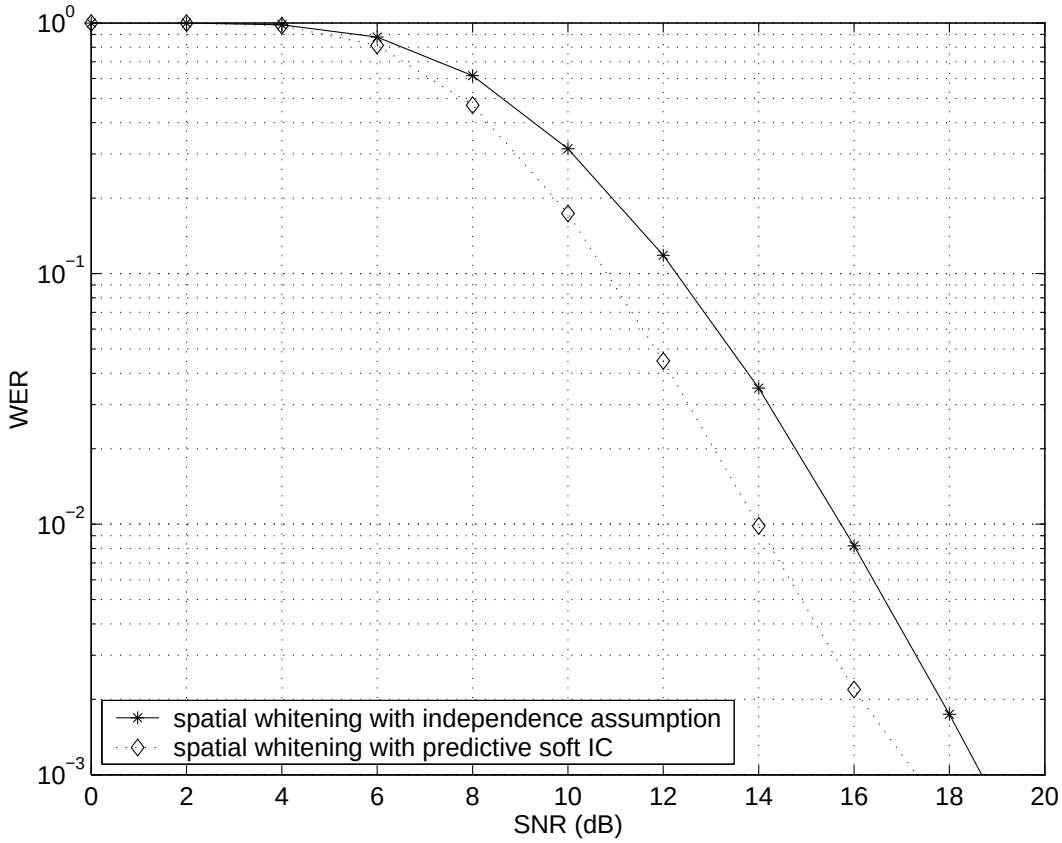


Figure 17: WER's for 2-space-time codes with 8 states, 4PSK.

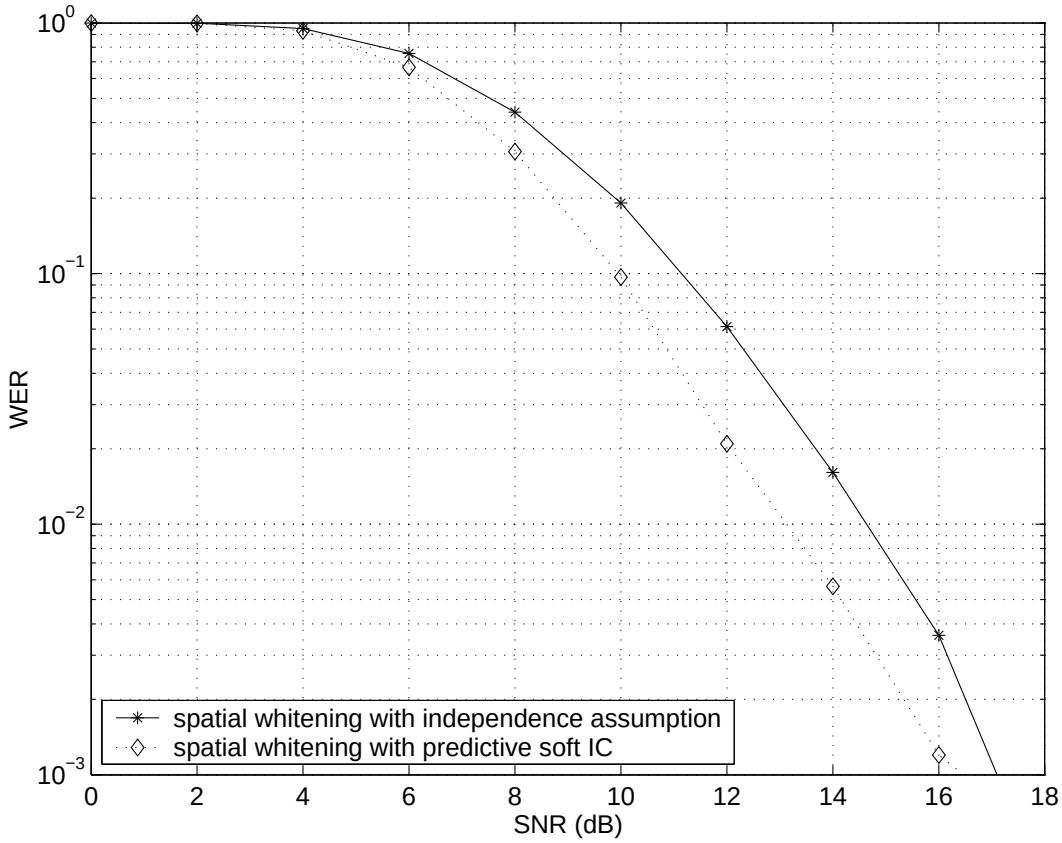


Figure 18: WER's for 2-space-time codes with 16 states, 4PSK.

reduced by interference cancellation, which implies more diversity can be gained for the desired component code. In addition, all component codes are simultaneously decoded so that coding delay is avoided. This is attractive when many component space-time codes are used and the coding delay becomes considerable.

CHAPTER V

QUASI-BLOCK DIAGONAL LDPC SPACE-TIME CODES (QBD-LDPC)

5.1 Introduction of LDPC

For a given space-time structure, FEC is a significant factor in system design to achieve a reasonable tradeoff between performance and complexity. For SISO systems, near-capacity-achieving codes have gained prominence in recent years since the invention of turbo codes [43] and the rediscovery of LDPC codes [44,45]. LDPC codes, originally introduced by Gallager in [46] 40 years ago, are a type of linear block codes with sparse parity-check matrices. Advanced coding schemes like Turbo codes and LDPC codes are both iteratively decoded in an effort to converge to the ML solution. The soft decoding of LDPC codes, based on the sum-product algorithm [47], is especially suited for VLSI implementation due to its high parallelizeability and thus yields smaller decoding delay compared with BCJR algorithm [48] for Turbo decoding. It has been shown that for *additive white Gaussian noise* (AWGN) channels, a properly constructed LDPC code is able to yield a threshold SNR within 0.0045 dB of the Shannon limit and simulation results demonstrate that it is within 0.04 dB of the Shannon limit at a BER of 10^{-6} with block length 10^7 [49].

LDPC codes are a special group of linear block codes. For any linear block code, each valid code word $\mathbf{b}$ satisfies the parity-check equations, that can be expressed in matrix form as

$$\mathbf{Eb} = \mathbf{0},$$

with the addition and multiplication performed in some Galois field. The parity-check

relations can be represented as a graph called Tanner graph, which is a bipartite graph with variable nodes on one side and parity-check nodes on the other. Each variable node represents a bit, and each parity-check node represents a parity-check constraint corresponding to a row of the parity-check matrix **E**. Bits, or variable nodes, involved in each parity-check constraint are connected to the corresponding parity-check node. Any linear code can be decoded with the help of this graphical representation using sum-product algorithm, whose convergence to the actual aposteriori probabilities is guaranteed when there are no cycles in the graph [50].

The ingenuity of LDPC codes is that they are defined by the parity-check matricesH, which has very few nonzero entries so that the hardware complexity in realizing the decoder is acceptable and can be constructed explicitly to avoid small cycles to improve convergence behavior. Some basic construction strategies of LDPC codes have been introduced in [51].

LDPC codes have been first introduced as regular graphs, which is called regular LDPC codes. However, it has been shown in [52] that irregular LDPC codes, based on irregular graphs, can indeed outperform regular ones. LDPC codes also have a threshold phenomenon. For a reasonable LDPC code with infinite block length, BER drops quickly with the increase of SNR when it is above a threshold, while the BER is large and has no significant change with SNR when it is below the threshold.

To understand the above phenomenon and design good LDPC codes, two prominent approaches have been developed. One approach is called density evolution [53], which is based on the concentration of the performance of the iterative decoder around its average performance. For LDPC codes with infinite length, the corresponding Tanner graph has no cycles. Therefore, the average behavior can be described by iteratively calculating message densities. The density evolution approach is further simplified in [54] by approximating message densities with Gaussian distributions.

Thus only means and variances of Gaussian densities need to be updated per iteration. The thresholds of LDPC codes can be accurately predicted by density evolution. For LDPC codes with an infinite block length, the key parameters to be optimized to approach Shannon capacity are the degree distributions of variable and parity-check nodes. The goal is to guarantee the convergence of message error rate to zero while maximizing the rate of the code at a given SNR or minimizing the threshold SNR given the code rate. Irregular LDPC codes are optimized based on density evolution in [55, 56] and density evolution with Gaussian approximation is used in [49].

The other popular approach is based on *extrinsic information transfer* (EXIT) chart, which has been first used by ten Brink to analyze turbo codes [57]. An EXIT chart plots the average extrinsic information coming out of the decoder as a function of the average a priori information going into the decoder. Several interesting properties of EXIT charts have been discussed in [58, 59]. It has shown from those properties that the design of capacity-approaching codes is essentially equivalent to a curve-fitting problem.

Due to its great success in SISO channels, LDPC codes are being considered as a coding scheme for wireless systems, especially MIMO systems. The general structure and decoding of space-time LDPC codes are introduced in [60]. Density evolution is used to study the performance and optimization of LDPC codes on flat Rayleigh fading channels in [61]. For frequency-selective channels, LDPC-coded MIMO-OFDM systems are investigated in [62,63], where density evolution is used to optimize LDPC codes for MIMO-OFDM systems. In [64], space-time LDPC codes are optimized for multicarrier systems with fixed channel response in frequency-selective environments. The challenge with iterative decoding in a general MIMO system is the extraction of a posteriori probabilities from the observed signal vector, which is the superposition of all transmitted signals. The derivation of the aposteriori probabilities requires an exhaustive search of all possible signal combinations. List (sphere) decoding may

be used to dramatically reduce the complexity. Still, a large list size is needed to achieve good performance for systems with high-order modulation [65]. Therefore, iterative coding combined with linear processing to separate signals is a promising performance-complexity tradeoff.

5.2 Quasi-Block Diagonal LDPC Space-Time Codes

LDPC can be directly applied to layered systems as the one-dimensional error correction code. The problem with layered systems is the presence of error-propagation, and the layers to be first detected, which usually have low SNR due to loss of signal power by linear nulling, are more likely to be incorrectly decoded. Therefore, optimum detection order can make a big difference for flat-fading channels. However, when implemented in frequency-domain, the advantage of detection order is less obvious for frequency-selective channels since the signals transmitted from the same antenna will experience different fadings across the frequency.

Therefore, we propose a space-time structure based on LDPC, which, similar to the layered space-time structure proposed in [6], can be detected with linear processing for efficient decoding. The lower-block triangular structure of the parity check matrix for the proposed LDPC code introduces correlation between consecutive layers so that the layers can be decoded successfully with the help of information from later layers.

An LDPC space-time code structure is devised where each layer comprises a sub-block and consecutive blocks are correlated by the introduction of *connection matrices* in the parity check matrix. The main novelty is that instead of demultiplexing the input data into separate streams and encoding each one independently, we are able to extract information from later layers to help improve the detection performance of the current layer, which is critical in preventing the error propagation in decision-feedback interference cancellation detectors.

The parity check matrix structure of binary QBD-LDPC for a MIMO system with

2 transmit and 2 receive antennas is illustrated in Figure 19. The entire matrix is denoted as $\mathbf{E}$ and any valid binary codeword $\mathbf{b}$ satisfies

$$\mathbf{Eb} = \mathbf{0}.$$

Each layer has the same length of codeword but the code rates may be different for different layers, which implies the numbers of the information bits may be different. In addition to the blocks along the main diagonal, which would be the parity check matrices for independent layers if all $\mathbf{E}_i^c = \mathbf{0}$, Sparse connection matrices, $\mathbf{E}_i^c$, link two consecutive layers i and $i+1$ as a channel for exchange of information. The layers are decoded from layer 1 to layer 2 and at detection stage i, the next layer $i+1$ will also contribute to the decoding of layer i.

As observed in [66], those bits (variable nodes) with higher degrees tend to converge faster, which motivates the design of irregular LDPC since the quickly converged bits will make it easier to decode the rest. Similarly, the introduction of connection matrices can be regarded as adding degrees to bits in the layer i so that those bits are better protected. In other words, when decoding layer i, $\mathbf{E}_i$, $\mathbf{E}_{i+1}$, and $\mathbf{E}_i^c$ form a smaller subcode where only the bits related to $\mathbf{E}_i$ with higher degrees are to be decoded at the current stage. The decoding of layer $i+1$ will be carried out later, with better channel quality after cancelling the interference from layer i, and with more protection, since layer $i+2$ will contribute to the decoding.

The system diagram of a QBD-LDBC based MIMO OFDM system is shown in Figure 20. Note that the output corresponding to each layer is permutated so that different parts of a layer are sent via different transmit antennas. The permutation is to guarantee all layers, on the average, have similar channel condition.

Next, we are going to introduce encoding and decoding of the proposed space-time structure.

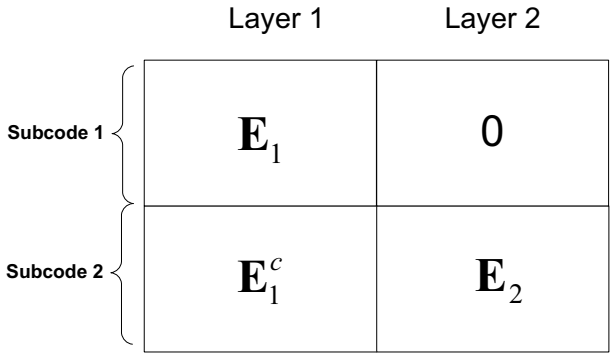


Figure 19: Parity-check matrix structure for QBD-LDPC space-time codes

5.3 Encoding

Let N be the length of the codeword for every layer. Assume the number of parity check bits for layer i is r_i, and the $(N - r_i) \times 1$ vector of input information bits is denoted as $\mathbf{u}_i$. The encoding of layer 1 is straightforward. By Gaussian elimination, we have

$$\mathbf{W}_1\mathbf{E}_1 = (\mathbf{P}_1 \ \mathbf{I}_{r_1}),$$

where $\mathbf{W}_1$ is a $r_1 \times r_1$ full rank matrix carrying out Gaussian elimination on $\mathbf{E}_1$, $\mathbf{P}_1$ is a $r_1 \times (N - r_1)$ matrix and $\mathbf{I}_{r_1}$ is an $r_1 \times r_1$ identity matrix. Then the codeword for layer 1 is formed by $\mathbf{b}_1 = \left(\mathbf{u}_1^T \ (\mathbf{P}_1\mathbf{u}_1)^T\right)^T$.

For layer i ($i > 1$), performing Gaussian elimination, we have

$$\mathbf{W}_i\mathbf{E}_i = (\mathbf{P}_i \ \mathbf{I}_{r_i}),$$

and the codeword for layer i is formed by

$$\mathbf{b}_i = \left(\mathbf{u}_i^T \ \left(\mathbf{P}_i\mathbf{u}_i + \mathbf{W}_i\mathbf{E}_{i-1}^c\mathbf{b}_{i-1}\right)^T\right)^T.$$

It is clear from the encoding processing that with a nonzero $\mathbf{E}_{i-1}^c$, part of the information about layer $i - 1$ is injected into the codeword of layer i.

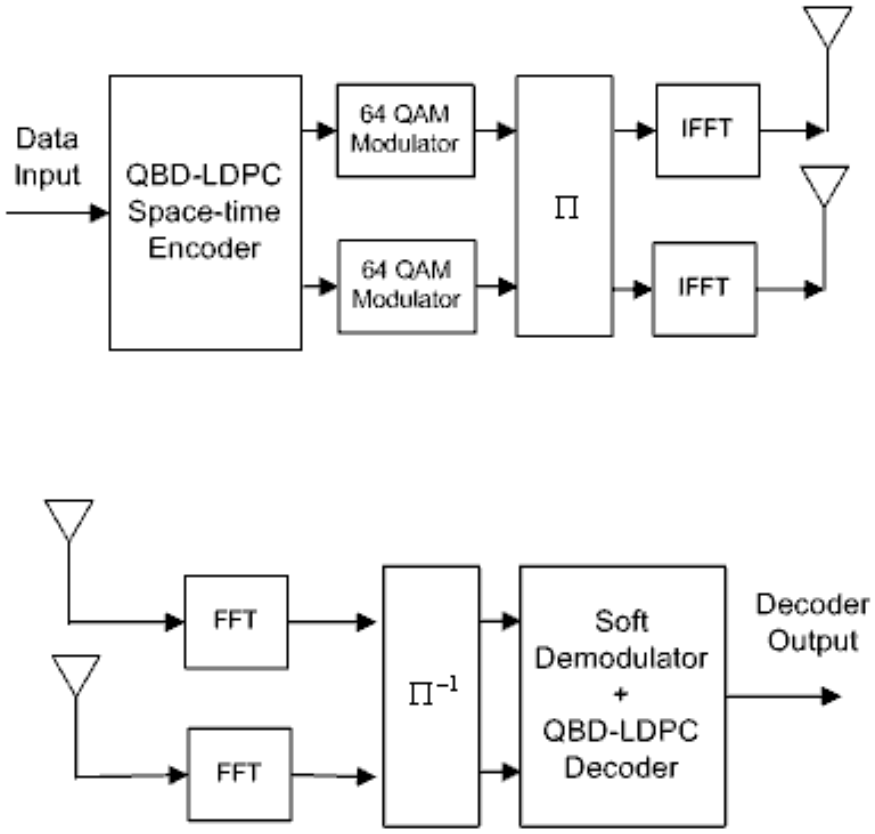


Figure 20: System diagram of quasi-block diagonal LDPC space-time codes.

5.4 Decoding

Assume we are now decoding layer i. Note that decoders for layer i and $i+1$ are both active. Again, we drop the subcarrier index k in Equation (8). Without loss of generality, assume the ith element of $\mathbf{r}[n]$ denoted as $r_i[n]$ is the signal from the ith layer, corresponding to the ith column of $\mathbf{H}[n]$, denoted as $\mathbf{H}_i[n]$.

Zero-Forcing

$$z_l[n] = \mathbf{w}_l^H[n]\,\mathbf{r}[n],\ l = i,\ i+1,$$

where the $N_r \times 1$ unit-norm weight vector $\mathbf{w}_j[n]$ nulls signals from all other undecoded layers according to the zero-forcing criterion.

Interference Cancellation

$$\tilde{z}_l[n] = z_l[n] - \sum_{m<l} \mathbf{w}_l^H[n]\,\mathbf{H}_m[n]\,\hat{s}_m[n],\ l = i,\ i+1,$$

where $\hat{s}_m[n]$'s are the reconstructed signals of decoded layers.

LDPC Decoding (Belief Propagation)

After zero-forcing and interference cancellation, the LDPC code is decoded like a 1-dimensional code at each stage. The *log-likelihood ratio* (LLR) is defined as

$$L(a) \triangleq \ln\left(\frac{p(a=1)}{p(a=0)}\right).$$

Then the soft information from the demodulator is

$$L_{zb}(b_k) = \ln \frac{\sum_{\mathbf{b}[n]:b_k=0} p(z|\mathbf{b}[n])\, e^{\sum_{l\in V_k} L_{bz}(b_l)}}{\sum_{\mathbf{b}[n]:b_k=1} p(z|\mathbf{b}[n])\, e^{\sum_{l\in V_k} L_{bz}(b_l)}},\ j = i,\ i+1, \tag{30}$$

where

$$p(z|\mathbf{b}[n]) = \frac{1}{\pi N_0} e^{-\frac{1}{N_0}|\tilde{z}_j[n] - \mathbf{w}_j^H[n]\cdot\mathbf{H}_j[n]s_j[n]|^2},\ j = i,\ i+1,$$

$$V_k = \{l | l \neq k \text{ and } b_l = 1\},$$

and $\mathbf{b}[n]$ is the bit vector containing b_k that is mapped to signal constellation $s_j[n]$.

The soft output from the demodulator is then sent to the sum-product decoder.

The message update at the variable node is

$$L_{bc}(b_k,\ c_l) = L_{zb}(b_k) + \sum_{\substack{c_j\in\Omega(b_k)\\ c_j\neq c_l}} L_{cb}(b_k,\ c_j). \tag{31}$$

and the message update at the check node is

$$L_{cb}(b_k,\ c_l) = L\left(\sum_{b_j\in\Omega(c_l)\backslash b_k} b_j\right), \tag{32}$$

which can be implemented efficiently by forward-backward algorithm using the property that

$$L(a+b) = \ln \frac{e^{L(a)} + e^{L(b)}}{1 + e^{L(a)+L(b)}},$$

or by the algorithm in [67]. Note that the message passing is performed between layer i and $i+1$, as well as within each layer.

Then the message passed to soft demodulator as a priori information is

$$L_{bz}(b_k) = \sum_{c_l \in \Omega(b_k)} L_{cb}(b_k,\ c_l), \tag{33}$$

where $\Omega(b_k)$ is the set of parity checks that involve b_k. The LLR for tentative decision is

$$LLR(b_k) = L_{zb}(b_k) + \sum_{c_l \in \Omega(b_k)} L_{cb}(b_k,\ c_l) = L_{zb}(b_k) + L_{bz}(b_k).$$

Note that one characteristic of LDPC is that the iteration can be terminated as soon as all parity checks are satisfied. In our case, we terminate the iteration when those parity checks involving only detected layers and the layer to be detected are satisfied. Take layer 1, for example. We stop the belief propagation when

$$\mathbf{E}_1 \mathbf{b}_1 = \mathbf{0},$$

is satisfied and the parity checks in

$$\mathbf{E}_1^c \mathbf{b}_1 + \mathbf{E}_2 \mathbf{b}_2 = \mathbf{0},$$

are only used for exchange of information between undetected layers to improve the detection performance.

After the current layer has been successfully decoded, the LLR's of the bits in the layer is set to $+\infty$ or $-\infty$ depending on their values to avoid ambiguity that may cause performance loss in the detection of later layers.

The structure of the decoder for QBD-LPDC coded layered systems is shown in Figure 21.

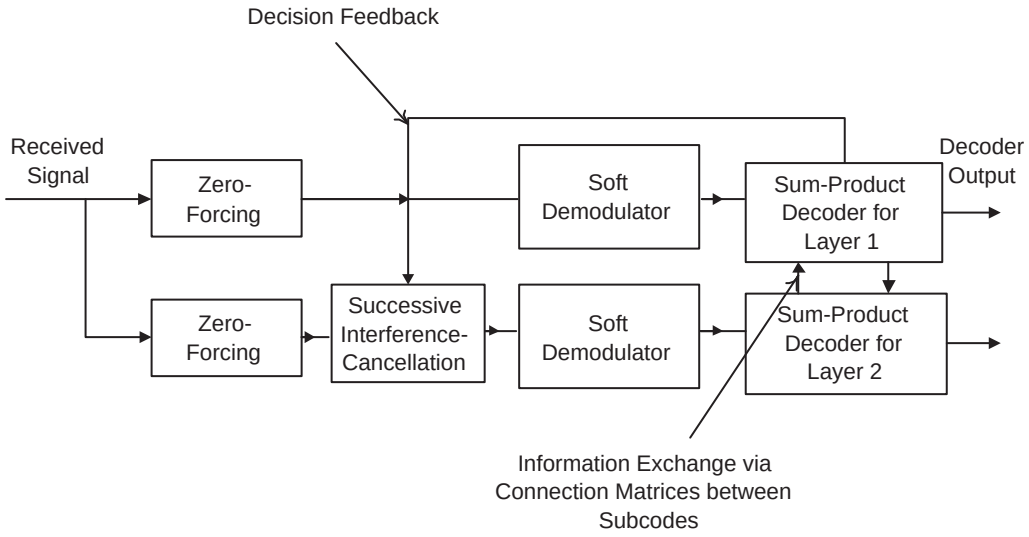


Figure 21: Decoder structure for QBD-LDPC codes in layered systems

5.5 Code Performance Analysis

5.5.1 Density Evolution

In this section we analyze the density evolution behavior of the QBD-LDPC. It should be noted that due to the particular structure and code rate distribution among layers, the density of non-zero entries in $\mathbf{E}_{i-1}$,$\mathbf{E}_{i-1}^{c}$ and $\mathbf{E}_i$ can not be the same, thus the classic density evolution approach needs to be modified to reflect this property. The degree distributions of variable nodes for $\mathbf{E}_{i-1}$ and $\mathbf{E}_i$ are specified by λ_l's and λ_l''s, respectively, where $\lambda_l(\lambda_l')$ is the fraction of edges connected to a variable node of local degree l inside $\mathbf{E}_{i-1}(\mathbf{E}_i)$. Equivalently, the degree distribution can be given by $\tilde{\lambda}_l(\tilde{\lambda}_l')$, which is the fraction of nodes of local degree l inside $\mathbf{E}_{i-1}(\mathbf{E}_i)$. Let the check degrees for $\mathbf{E}_{i-1}$, $\mathbf{E}_{i-1}^{c}$ and $\mathbf{E}_i$ be as regular as possible, as suggested by the highly concentrated distribution of check degrees for good irregular codes [68]. Each column of $\mathbf{E}_{i-1}^{c}$ has exactly κ non-zero entries.

Symmetry of LLR distribution is an important property, which is invariant under density evolution. It has been proved in [69] that the initial LLR density function in Rayleigh fading channels also satisfies the symmetry property if *channel state*

information (CSI) is known at the receiver. However, the LLR density function of the output of the soft demodulator has no closed-form expression. Following [70], we approximate the distribution as a mixture of a number of Gaussian distributions.

$$p_{zb} \sim \sum_{t=1}^{T} \alpha_t N\left(\mu_t\right),$$

where $N\left(m\right)$ represents a Gaussian probability density function with mean μ and variance 2μ. The parameters involved are estimated by *expectation maximization* (EM) algorithm from the LLR samples taken for each layer after zero-forcing and perfect interference cancellation.

Next we formulate the message passing as information exchange between connected blocks. Let b denote the bits of layer $i-1$, b' be the bits of layer i, c be the parity checks only containing bits from layer $i-1$, and c' be the parity checks containing bits from layer i. We ignore the changes caused by a priori information on the output density function of the soft demodulator, assuming the Gray-type modulation makes bits completely independent. We also assume the LLR's of bits in decoded layers already converge to infinity so that we only have to consider the convergence of two consecutive layers.

$b \rightarrow$ c:

For a variable bit of layer $i-1$ with degree l inside $\mathbf{E}_{i-1}$, the density functions of message passed to check nodes are

$$p_{bc,l} = \sum_{t=1}^{T} \alpha_t N\left(\mu_t + \left(l-1\right) m_{cb} + \kappa m_{c'b}\right),$$

and

$$p_{bc',l} = \sum_{t=1}^{T} \alpha_t N\left(\mu_t + i m_{cb} + \left(\kappa - 1\right) m_{c'b}\right).$$

The message density function from a variable bit of layer i with degree l inside $\mathbf{E}_i$ to check nodes is

$$p_{b'c',l} = \sum_{t=1}^{T} \alpha_t N\left(\mu_t + \left(l-1\right) m_{c'b'}\right),$$

$c \longrightarrow b$:

Let $\phi(x) \triangleq E\{\tanh(N(x)/2)\}$,

$$m_{bc,l,t} \triangleq \mu_t + (l-1)\, m_{cb} + \kappa m_{c'b}$$

$$m_{bc',l,t} \triangleq \mu_t + i m_{cb} + (\kappa - 1)\, m_{c'b}$$

and

$$m_{b'c',l,t} \triangleq \mu_t + (l-1)\, m_{c'b'}$$

Assume the LLR at the output of check nodes is Gaussian, the means of the distribution satisfy

$$\phi(m_{cb,j}) = \left(\sum_{t=1}^{T} \sum_{l=1}^{d_{v,max}} \alpha_t \lambda_l \phi(m_{bc,l,j}) \right)^{j-1},$$

$$\begin{aligned} \phi(m_{c'b,j_1,j_2}) &= \left(\sum_{t=1}^{T} \sum_{l=1}^{d_{v,max}} \alpha_t \tilde{\lambda}_l \phi(m_{bc',l,t}) \right)^{j_1-1} \\ &\quad \cdot \left(\sum_{t=1}^{T} \sum_{l=1}^{d'_{v,max}} \alpha_t \lambda'_l \phi(m_{b'c',l,t}) \right)^{j_2}, \end{aligned}$$

and

$$\begin{aligned} \phi(m_{c'b',j_1,j_2}) &= \left(\sum_{t=1}^{T} \sum_{l=1}^{d_{v,max}} \alpha_t \tilde{\lambda}_l \phi(m_{bc',l,t}) \right)^{j_1} \\ &\quad \cdot \left(\sum_{t=1}^{T} \sum_{l=1}^{d'_{v,max}} \alpha_t \lambda'_l \phi(m_{b'c',l,t}) \right)^{j_2-1}, \end{aligned}$$

respectively. $d_{v,max}(d_{c,max})$ is the maximum degree for variable nodes(check nodes).

We locally combine different distributions, within $\mathbf{E}_{i-1}$,$\mathbf{E}^c_{i-1}$ and $\mathbf{E}_i$, respectively, to yield the LLR density function passed from check nodes to variable nodes due to the non-uniform density of the code. Define

$$\hat{\rho}_{j_1,j_2} \triangleq \frac{\tilde{\rho}^c_{j_1} \tilde{\rho}'_{j_2} j_1}{\sum_{j_1=2}^{d^c_{c,max}} \sum_{j_2=2}^{d'_{c,max}} \tilde{\rho}^c_{j_1} \tilde{\rho}'_{j_2} j_1},$$

and

$$\hat{\rho}'_{j_1,j_2} \triangleq \frac{\tilde{\rho}^c_{j_1}\tilde{\rho}'_{j_2}j_2}{\sum_{j_1=2}^{d^c_{c,max}}\sum_{j_2=2}^{d'_{c,max}}\tilde{\rho}^c_{j_1}\tilde{\rho}'_{j_2}j_2}.$$

If the distribution of number of edges per parity check in $\mathbf{E}^c_{i-1}$ is independent of that in $\mathbf{E}_i$, the density functions become

$$p_{cb} = \sum_{j=2}^{d_{c,max}} \rho_j N(m_{cb,j}),$$

$$p_{c'b} = \sum_{j_1=2}^{d^c_{c,max}}\sum_{j_2=2}^{d'_{c,max}} \hat{\rho}_{j_1,j_2} N\left(m_{c'b,j_1,j_2}\right),$$

and

$$p_{c'b'} = \sum_{j_1=2}^{d^c_{c,max}}\sum_{j_2=2}^{d'_{c,max}} \hat{\rho}'_{j_1,j_2} N\left(m_{c'b',j_1,j_2}\right).$$

The threshold SNR is defined as the minimum SNR at which an LDPC converges or, equivalently under Gaussian approximation, the mean of the message density goes to infinity. Note that in a successive decoding scheme, there are N_t stages where the convergence of each corresponding part in the code has to be considered.

5.5.2 An Example

In this subsection, we give an example to explain the convergence behavior of QBD-LDPC code. The LLR samples are computed for each layer assuming perfect detection of previous layers and thus perfect successive interference cancellation. Then the LLR samples are used to derive the mixture of Gaussian distributions to approximate the exact LLR distribution, as stated in the previous subsection. To successfully decode the entire codeword, each layer has to converge at each detection stage and the threshold SNR is the maximum of the threshold SNR's for all detection stages as determined by the process described in the previous subsection. We select a 4-layer QBD-LDPC code with all E_i's having the same structure, hence the same degree distribution. The variable degree distribution is given in Table 4.

Table 4: Degree Distribution of an example 4-layer QBD-LDPC code, λ_{max}=5.

λ_2	λ_3	λ_4	λ_5
0.4188	0.3581	0.2145	0.0086

The threshold SNR's for each layer with various κ's are given in Table 5. Note that all SNR's are the channel SNR instead of post-processing SINR at each detection stage. It is seen from the table that due to the bad channel quality when detecting the first layer, the threshold channel SNR for the first layer to converge is greater than 20 dB for independently LPDC-encoded layered structure. With the presence of the connection matrix, the threshold SNR is reduced to 17.86 dB and 17.32 dB, with κ being one and two, respectively. We further observe that the threshold SNR's no longer decrease for κ's beyond 3, which suggests that it is sufficient to keep a sparse structure for the connection matrix.

When comparing the threshold SNR's of layer 3, we see that the introduction of the connection matrix does not improve the convergence behavior. From the table, layer 4 has a higher threshold SNR after interference from all layers is perfectly subtracted, which means it has an even higher threshold SNR at the time layer 3 is being decoded. Thus, the correlation between layer 3 and 4 will not be able to help layer 3 converge faster.

Table 5: Threshold SNR's.

κ	Layer 1	Layer 2	Layer 3	Layer 4
0	$>$ 20 dB	15.88 dB	15.07 dB	15.43 dB
1	17.86 dB	15.07 dB	15.07 dB	15.43 dB
2	17.32 dB	15.07 dB	15.07 dB	15.43 dB
3	17.32 dB	15.07 dB	15.07 dB	15.43 dB
4	17.32 dB	15.07 dB	15.07 dB	15.43 dB
5	17.32 dB	15.07 dB	15.07 dB	15.43 dB

5.6 Simulation Results

In this section we simulate the performance of a MIMO-OFDM system with QBD-LDPC structure. The simulated system has 2 transmit and 2 receive antennas with 64QAM modulation. IEEE 802.11 TGn channel model "F" is used. The Matlab program provided by [71] is used to generate channel parameters. The structure of a OFDM symbol conforms to the IEEE 802.11a PHY standard in the 5GHz band. 48 out of 64 subcarriers are used for data transmission in a 20MHz system bandwidth and the guard interval is 0.8μs, resulting in a total of 4μs symbol duration. Perfect channel estimation is assumed and the channel is assumed to be constant during each transmission. The transmission of each layer is cycled through all transmit antennas. The codeword length of each layer is 1152 bit corresponding to 4 OFDM symbols and the code rate for each layer is 1/2. Therefore, the data rate of the simulated system is 72Mbps.

The QBD-LDPC with the specified parameters is generated by the modified version of progressive edge growth construction algorithm [72]. $\mathbf{E}_1$ and $\mathbf{E}_2$ are first generated progressively to maximize the global girth of the corresponding Tanner graph. Then the number of bits involved in each parity check equation in $\mathbf{E}_1^c$ as in Figure 19 is set to 2, introducing connection between the two layers. The addition of each edge also tries to maximize the girth.

The WER of the proposed system is given in Figure 22. The WER of a system with independent LDPC-coded layers (V-BLAST) is provided for comparison. For a WER of 1%, the required SNR for the proposed system is 19.6dB, which is about 0.5dB better than the V-BLAST system.

5.7 Summary

In this chapter, we introduce a layered space-time coding that is characterized by the low-block triangular structure of the LDPC parity-check matrix. The layers

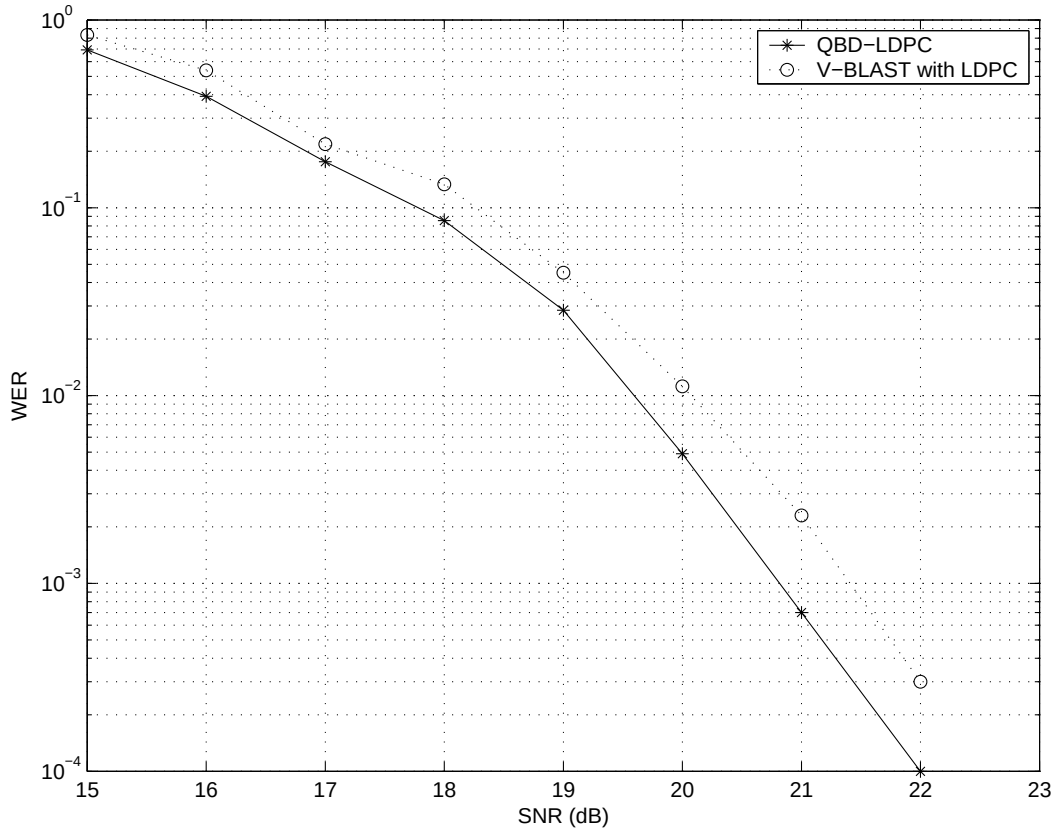


Figure 22: WER's of LDPC-coded layered systems

are detected successively, each with the help of information from the next layer to enhance performance. Encoding and decoding algorithm are introduced and the simulation shows a 0.5dB improvement over the conventional LDPC-coded V-BLAST at WER=1%.

CHAPTER VI

STATISTICAL RATE ALLOCATION

The low complexity of layer processing makes the layered structure a promising candidate for systems with a large number of transmit antennas and higher order modulation. However, in the original layered system, the input data is evenly divided into substreams and all layers have the same code rate. Due to the loss of signal energy and degree of freedom by nulling, the channel quality for the layers to be first detected frequently can not support reliable transmission with the given layer data rate and those layers are more error-prone. To remedy this problem, optimum detection ordering has been proposed [73] [74]. The idea is to always select the layer with the best channel quality among the remaining layers to be detected first. The drawback of detection ordering is the high computational complexity, especially in frequency-selective channels, such as wide-band OFDM systems, where the channel responses are different for different subcarriers. Another issue is that the channel quality difference between different layers tends to decrease with increasing frequency-selectivity, making the gain of using optimum detection order less significant. Therefore, the original V-BLAST system can only achieve a portion of the system capacity, even with optimum detection ordering.

Based on the observation above, we suggest that fixing the detection order while adapting the data rate for each layer depending on the detection order is a more promising solution for frequency-selective channels. Intuitively, the data rate should increase for later layers corresponding to the increasing channel quality for later detection stages. Rate adaptation and/or power allocation for layered structure depending on the instantaneous post-processing SINR have been investigated in [75] [76] [77].

An interesting fact has been discovered that with MMSE interference nulling and properly selected rate for each layer, the sum of capacities of all layers (with perfect SIC) is exactly the instantaneous open-loop capacity [78] [79].

To achieve the open-loop capacity, instantaneous rate feedback is needed, which introduces transmission overhead. In addition, for fast-varying environments, up-to-date rate feedback may not be feasible. However, for channels with high frequency selectivity or enough time variation, the variations of layer capacities tend to decrease, as suggested by the law of large numbers. Thus, we can approach the open-loop channel capacity by determining the rate for each layer based on the knowledge of the channel statistics, but not of the actual realizations. Our approach is to minimize the overall layer outage probability given the total information rate. The key difference of our work from that in [75] [76] [77] is that we capitalize on the statistical instead of instantaneous information of the channel so that frequent rate feedback is not necessary.

6.1 Rate Allocation for Layered Systems

6.1.1 Statistical Rate Allocation

In the proposed system, the detection order is fixed and the data rate for each layer is adjusted according to the channel quality and detection order of that particular layer. Without loss of generality assume the order of detection is from transmit antenna 1 to N_t. The instantaneous information rate to be allocated to transmit antenna l is [78] [79] [80]

$$\begin{aligned} C_l &= \log_2 \det\left(\mathbf{I}_{N_r} + \frac{SNR}{N_t}\mathbf{H}_{(l-1)}\mathbf{H}_{(l-1)}^H\right) \\ &\quad - \log_2 \det\left(\mathbf{I}_{N_r} + \frac{SNR}{N_t}\mathbf{H}_{(l)}\mathbf{H}_{(l)}^H\right), \end{aligned} \tag{34}$$

where $\mathbf{H}_{(l)} = [\mathbf{h}_{l+1}\ \mathbf{h}_{l+2}\ \cdots\ \mathbf{h}_{N_t}]$, and $\mathbf{h}_l$ is the lth column of $\mathbf{H}$. It is obvious that $\mathbf{H}_{(0)} = \mathbf{H}$.

It has been observed by various researchers that the distribution of the capacity of a MIMO channel, Rayleigh or Rician, can be accurately approximated by a Gaussian distribution at medium and high SNR's [81] [82]. Thus the instantaneous capacity of each layer is also Gaussian distributed, denoted as

$$C_l \sim N\left(\eta_l, \sigma_l^2\right),$$

where η_l and σ_l^2 are the mean and variance of the capacity of the l-th layer, respectively.

To obtain the optimum rate allocation, the mean and variance of layer capacities are needed. For flat-fading i.i.d complex Gaussian channel, the mean may be derived directly from the results by Telatar [1], which is an integral involving Laguerre polynomials, and no simple explicit expression is available. Simple approximate formulas by using asymptotic results in [83] [84] may be used as an alternative. However, explicit expressions for a general channel model with spatial correlation are hard to derive, especially for the variance in frequency-selective channels. Therefore, we just compute the mean and variance by taking enough samples of the channel parameters, which is a feasible assumption since the system can always get this information from past channel observations.

In order to properly select data rates for different layers, we minimize P_{out}, the outage probability of the whole layered system. It is equivalent to maximizing the probability when no layers have information rates greater than the respective layer capacities, i.e.,

$$1 - P_{out} = \prod_{l=1}^{N_t}(1 - P_l) = \prod_{l=1}^{N_t} \int_{u_l}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_l} e^{-\frac{(t-\eta_l)^2}{2\sigma_l^2}} dt,$$

subject to the constraint that the total information rate is fixed,

$$\sum_{l=1}^{N_t} u_l = C_T,$$

where u_l is the information rate allocated to the l-th layer. P_l is the outage probability for the l-th layer.

Denote

$$u_l = x_l + \eta_l.$$

By setting up the equivalent Lagrangean, we try to find stationaries points, i.e.,

$$\begin{aligned} J &= \ln\left(\prod_{l=1}^{N_t}\int_{x_l}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_l}e^{-\frac{t^2}{2\sigma_l^2}}\,dt\right) \\ &\quad -\lambda\left(\sum_{l=1}^{N_t}x_l+\sum_{l=1}^{N_t}\eta_l - C_T\right). \end{aligned} \tag{35}$$

It can be easily verified the stationary point satisfies

$$-\frac{\frac{1}{\sqrt{2\pi}\sigma_l}e^{-\frac{x_l^2}{2\sigma_l^2}}}{\frac{1}{\sqrt{2\pi}\sigma_l}\int_{x_l}^{\infty}e^{-\frac{t^2}{2\sigma_l^2}}\,dt} = \lambda,$$

for $l = 1,\ 2,\ \cdots,\ N_t$. Since we are only interested in the case when the input data rate is set such that reliable transmission is guaranteed most of the time, we assume that the outage probability given the total information rate is small, or equivalently, $x_l < 0$, and $|x_l| \gg \sigma_l$. Thus,

$$\frac{1}{\sqrt{2\pi}\sigma_l}\int_{x_l}^{\infty}e^{-\frac{t^2}{2\sigma_l^2}}\,dt \approx 1,$$

and

$$\frac{x_l^2}{\sigma_l^2} \approx 2\ln\sigma_l + \ln 2\pi\lambda^2, \tag{36}$$

where the left hand side of Equation (36) is approximately the same for all layers, since it is observed that σ_l's have magnitudes of the same order. We then have

$$x_l^* \approx \frac{\sigma_l}{\sum_{m=1}^{N_t}\sigma_m}\left(C_T - \sum_{m=1}^{N_t}\eta_m\right),$$

and the optimum rate for the l-th layer is

$$u_l^* \approx \eta_l + \frac{\sigma_l}{\sum_{m=1}^{N_t}\sigma_m}\left(C_T - \sum_{m=1}^{N_t}\eta_m\right). \tag{37}$$

Therefore, the optimum outage probability for each layer is

$$\begin{aligned} P_l^* &= \int_{-\infty}^{x_l^*}\frac{1}{\sqrt{2\pi}\sigma_l}e^{-\frac{t^2}{2\sigma_l^2}}\,dt \\ &= \int_{-\infty}^{\frac{C_T-\sum_{m=1}^{N_t}\eta_m}{\sum_{m=1}^{N_t}\sigma_m}}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt, \end{aligned}$$

which is the same for all layers. Thus, the optimum data rate allocation is that the data rate for each layer is backed off from the mean capacity of the layer, and the amount of backoff is proportional to the standard deviation of the layer capacity, which agrees with our intuition that more backoff is needed for the more fluctuant channel. In addition, the minimum overall layer outage probability is achieved when each layer has the same layer outage probability. Define the normalized capacity margin as

$$\varphi \triangleq \frac{\sum_{m=1}^{N_t} \eta_m - C_T}{\sum_{m=1}^{N_t} \sigma_m}.$$

The optimum overall layer outage probability is then

$$\begin{aligned} P_{out}^* &= 1 - \prod_{l=1}^{N_t} (1 - P_l^*) \\ &= 1 - \left(\int_{-\varphi}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \right)^{N_t}, \end{aligned} \tag{38}$$

which states the interesting fact that the minimum overall layer outage probability of a layered system is uniquely determined by the normalized capacity margin.

It should be noted that the approach proposed above can be applied to cases where the association of transmit antennas with substreams varies, or the channel is frequency-selective such as OFDM systems. We only have to sum up all the data rates as given by Equation (34) for each substream and determine the corresponding mean and variance of the channel capacity for each substream.

6.1.2 Relationship Between Outage Probabilities

In this subsection, we derive some relationship between the outage probability of a layered system and that of the channel. Using the approximation [85],

$$\int_x^{\infty} e^{-\frac{t^2}{2}} dt \approx \frac{e^{-\frac{x^2}{2}}}{x}, \; x \gg 1, \tag{39}$$

then from Equation (38),

$$P_{out}^* \approx \frac{N_t}{\sqrt{2\pi}\varphi} e^{-\varphi^2/2}.$$

Similarly, we can derive the asymptotic outage probability of the MIMO channel with the same overall information rate C_T as

$$P_{ch} \approx \frac{1}{\sqrt{2\pi}\varphi_{ch}} e^{-\varphi_{ch}^2/2},$$

where

$$\varphi_{ch} = \frac{\eta_{ch} - C_T}{\sigma_{ch}},$$

η_{ch} is the ergodic MIMO channel capacity, and σ_{ch}^2 is the variance of the MIMO channel capacity. Note that

$$\eta_{ch} = \sum_{l=1}^{N_t} \eta_l,$$

and

$$\sigma_{ch} \leq \sum_{l=1}^{N_t} \sigma_l,$$

since

$$E\left\{\left(\sum_l v_l\right)^2\right\} \leq \left(\sum_l \sqrt{E\left\{v_l^2\right\}}\right)^2,$$

for any set of random variables v_l's.

Thus,

$$\varphi_{ch} \geq \varphi,$$

and

$$P_{out} \geq P_{out}^* \approx \frac{N_t}{\sqrt{2\pi}\varphi} e^{-\varphi^2/2} \geq N_t \frac{1}{\sqrt{2\pi}\varphi_{ch}} e^{-\varphi_{ch}^2/2} \approx N_t P_{ch},$$

which implies that with the same overall information rate, the asymptotic outage probability of layered structure is at least N_t times that of the MIMO channel. However, the low complexity of layered structure makes it still a good candidate for high-speed wireless systems.

6.2 Rate Selection for Maximum Effective Throughput

In the previous section, we have introduced our algorithm of optimum statistical data rate allocation for a given total data rate. In practice, we want to use the channel to

transmit as much information as possible. If the total data rate is too low, the total successfully transmitted information is limited by the total amount of information during each transmission even though each transmission is correctly detected. On the other hand, if the total data rate is set too high, almost every transmission is lost because of the high outage probability. Therefore, we expect a best total data rate at each SNR that maximizes the effective throughput. In this section, we will specify the condition that the best total data rate satisfies with Gaussian approximation of layer capacities.

Effective throughput is defined as the portion of input data rate that is transmitted successfully, and in our case, is the product of the total input data rate and the probability that the channel supports the total input data rate.

$$C_{eff} = C_T * (1 - P_{out}) = C_T \left(\int_{\frac{C_T - \sum_{m=1}^{N_t} \eta_m}{\sum_{m=1}^{N_t} \sigma_m}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \right)^{N_t}.$$

$$\frac{\partial C_{eff}}{\partial C_T} = \frac{\partial}{\partial C_T} \left\{ C_T \left(\int_{\frac{C_T - \sum_{m=1}^{N_t} \eta_m}{\sum_{m=1}^{N_t} \sigma_m}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \right)^{N_t} \right\} = 0.$$

It is easy to verify that the stationary point satisfies

$$\int_{\frac{C_T - \sum_{m=1}^{N_t} \eta_m}{\sum_{m=1}^{N_t} \sigma_m}}^{\infty} e^{-\frac{t^2}{2}} dt - \frac{N_t}{\sum_{m=1}^{N_t} \sigma_m} e^{-\frac{\left(\frac{C_T - \sum_{m=1}^{N_t} \eta_m}{\sum_{m=1}^{N_t} \sigma_m} \right)^2}{2}} = 0, \tag{40}$$

which is not easy to solve. Instead, we try to derive an asymptotic solution as $\sum_{m=1}^{N_t} \sigma_m$ tends to zero. This occurs when we have K independent channel realizations and K tends to infinity. Then the sum of the standard deviation becomes

$$\lim_{K \to \infty} \sum_{m=1}^{N_t} \sigma'_m[K] = \lim_{K \to \infty} \sum_{m=1}^{N_t} \frac{\sigma_m}{\sqrt{K}} = 0,$$

where $\sigma'^2_l[K]$ is the variance of the capacity of the l-th layer with K independent channel realizations.

Using the same approximation as in Equation (39), Equation (40) becomes

$$\begin{aligned}
&\int_{\frac{C_T-\sum_{m=1}^{N_t}\eta_m}{\sum_{m=1}^{N_t}\sigma_m}}^{\infty} e^{-\frac{t^2}{2}}dt - \frac{N_t}{\sum_{m=1}^{N_t}\sigma_m} e^{-\frac{\left(\frac{C_T-\sum_{m=1}^{N_t}\eta_m}{\sum_{m=1}^{N_t}\sigma_m}\right)^2}{2}} \\
\approx &\sqrt{2\pi} + e^{-\frac{\left(\frac{C_T-\sum_{m=1}^{N_t}\eta_m}{\sum_{m=1}^{N_t}\sigma_m}\right)^2}{2}} \left\{ \frac{\sum_{m=1}^{N_t}\sigma_m}{C_T-\sum_{m=1}^{N_t}\eta_m} - \frac{N_t}{\sum_{m=1}^{N_t}\sigma_m} \right\} \\
\approx &\sqrt{2\pi} - e^{-\frac{\left(\frac{C_T-\sum_{m=1}^{N_t}\eta_m}{\sum_{m=1}^{N_t}\sigma_m}\right)^2}{2}} \frac{N_t}{\sum_{m=1}^{N_t}\sigma_m} \\
= &0.
\end{aligned}$$

Therefore,

$$C_T^* \approx \sum_{m=1}^{N_t}\eta_m - \sum_{m=1}^{N_t}\sigma_m \sqrt{2\ln\frac{N_t}{\sqrt{2\pi}\sum_{m=1}^{N_t}\sigma_m}}. \tag{41}$$

To summarize, the rate adaption of the proposed system proceeds in two steps. First, the mean and standard deviation of layer capacities are used to determine the total information rate of the system by Equation (41). Then the data rate of each layer is determined by Equation (37), depending on the channel statistics and detection order of the layer. Note that some additional SNR margin is necessary in practical systems due to the fact that the coding is not capacity-approaching.

6.3 Simulation Results

First, we use the outage probability to evaluate the performance gap between layered structure and the MIMO channel. We compare the outage probability of a 2x2 layered OFDM system with the channel outage probability, which may serve as a lower bound for WER of practical systems. Each substream is cyclically associated with transmit antennas at different subcarriers to effectively reduce the variation of layer capacities. The OFDM symbol structure conforms to the IEEE 802.11a PHY standard with 20 MHz system bandwidth in the 5 GHz band. We consider the performance in frequency-selective channels with correlated fading at the antenna elements.

Specifically, the standard IEEE 802.11 TGn channel models 'B' and 'D' generated by the Matlab program available at [71] are used. Channel 'B' has a shorter delay spread than channel 'D', thus has less frequency-selectivity.

Figures 23.a and 23.b compare the outage probability of our proposed system and that of an equal-rate V-BLAST system. The total information rate per subcarrier is 9 bps/Hz. It can be seen that for an outage probability of 1%, the gap between outage probability of the channel and that of the proposed system reduces from 3.2 dB for channel 'B' to 1.3 dB for channel 'D'. It can be explained by the fact that increased frequency-selectivity reduces the capacity variation of each layer. The SNR improvement for an outage probability of 1% is 6.3dB and 3.6dB for channel 'B' and 'D', respectively.

To test the sensitivity of the optimum rate allocation to SNR, the outage probabilities of the rate allocation optimized for SNR=24 dB are compared with those where the rates are optimized separately for each SNR. The performance degradation of optimized rate allocation for a fixed SNR is small in the 2x2 system, suggesting the robustness of the rate allocation to SNR disturbance. In addition, optimum rate allocation derived by a brute force method is also investigated. In the brute force method, each grid point with each coordinate being an integer multiple of 0.01 bps/Hz is evaluated to find the optimum of Equation (35). The approximate solutions and the solutions derived by brute force method are further compared in figures 24.a and 24.b for channel 'B' and 'D', respectively. It can be seen that the approximate solution gives quite accurate results, while being much simpler to obtain.

Then, similar numerical results are obtained for a 4x4 system. The outage probabilities are shown in figures 25.a and 25.b, for channel 'B' and 'D', respectively. The total information rate per subcarrier is 18 bps/Hz. It is seen that there is an even bigger performance gap between the optimum rate allocation and the equal-rate

V-BLAST. The proposed system is 2.5 dB and 1.5 dB away from the channel outage capacity at a WER of 1%. Note in figure 25.a that the outage probability of the rate allocation optimized for SNR=24 dB is even smaller than that of optimum per-SNR allocation. This is probably due to the the Gaussian approximation of the distribution of channel capacities.

Next, we compare the performance of 2x2 layered systems, given the same or similar total input information rate. Perfect channel estimation and synchronization is assumed. Each packet contains 1k Bytes of data and 10,000 packets are sent for each SNR. In the conventional system, each substeam is coded using rate 3/4 industry standard convolutional coding, interleaved, and modulated using 64QAM constellation. Two substreams are then sent simultaneously, resulting in a total data rate of 108 Mbps.

Due to the limited number of supported data rates in a practical system, only BPSK, QPSK, 16QAM, and 64QAM with convolutional coding of code rate 1/2, 2/3, 3/4, and 7/8 may be used for transmission of each layer. Therefore, the optimum data rates has to be quantized to the closest supported transmission modes and the code rates, modulation schemes for different layers may be different depending on the statistics of layer channel quality after layered processing. The nominal total information rate as input to the rate allocation algorithm is set such that the resultant total information rate after rate allocation is as close to that of the conventional system as possible to make the comparison reasonable. The rate allocation is optimized for the lowest SNR such that the optimum overall layer outage probability is below 1%. The optimized rates are then shown in Table 6.a and Table 6.b, for channel 'B' and 'D', respectively. The total data rate is 99 Mbps and 111 Mbps, respectively, as compared to the 108 Mbps of the conventional layered system.

In Figure 26.a, the WER's of the proposed system and the conventional layered system are shown. For a WER of 10%, the proposed system provides an SNR gain

Table 6: Transmission modes of different layers in a 2x2 system with statistical rate allocation. (a) Channel 'B'. (b) Channel 'D'.

Detection order	Modulation	Code rate	Layer data rate (Mbps)
1	16QAM	3/4	36
2	64QAM	7/8	63

(a)

Detection order	Modulation	Code rate	Layer data rate (Mbps)
1	64QAM	2/3	48
2	64QAM	7/8	63

(b)

of about 5 dB and 3 dB for channel model 'B' and 'D', respectively. The effective throughput of the same system is shown in Figure 26.b. It is seen that with the same SNR, the proposed system yields a maximum increase of about 38 Mbps and 32 Mbps in effective throughput, for channel 'B' and 'D', respectively, while maintaining similar peak data rate.

Finally, a 4x4 system is simulated with results shown in Figure 27. For the equal-rate V-BLAST, each of the 4 substreams has the same configuration as the 2x2 case, with the total data rate being 216 Mbps. The optimum rate allocations that deliver similar data rate are shown in Table 7.a and Table 7.b, with the resultant total data rates being 221 Mbps and 212 Mbps, for channel 'B' and 'D', respectively. For a WER of 10%, the SNR gain of the proposed system is 4.5 dB and 6.65 dB for channel model 'B' and 'D', respectively. From Figure 27.b, the maximum increase in effective throughput is about 91 Mbps and 174.5 Mbps, for channel 'B' and 'D', respectively.

Table 7: Transmission modes of different layers in a 4x4 system with statistical rate allocation. (a) Channel 'B'. (b) Channel 'D'.

Detection order	Modulation	Code rate	Layer data rate (Mbps)
1	16QAM	2/3	32
1	64QAM	7/8	63
1	64QAM	7/8	63
2	64QAM	7/8	63

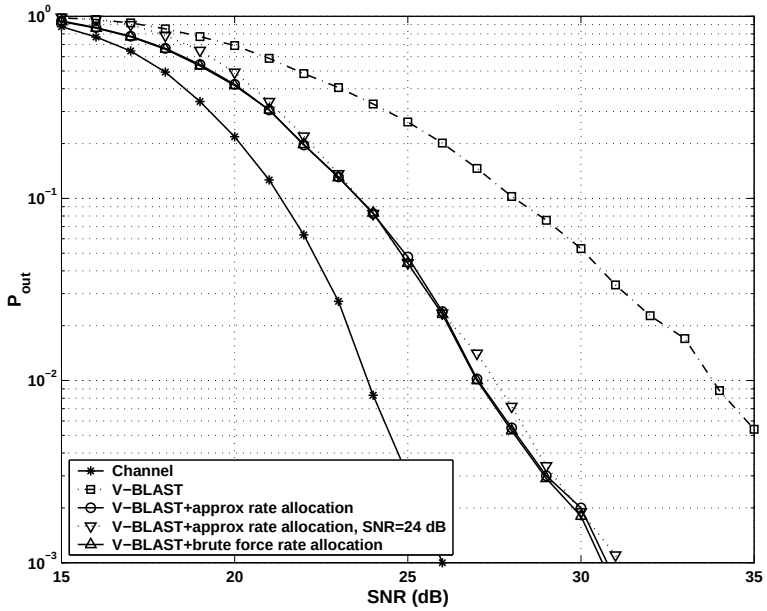
(a)

Detection order	Modulation	Code rate	Layer data rate (Mbps)
1	16QAM	2/3	32
1	64QAM	3/4	54
1	64QAM	7/8	63
2	64QAM	7/8	63

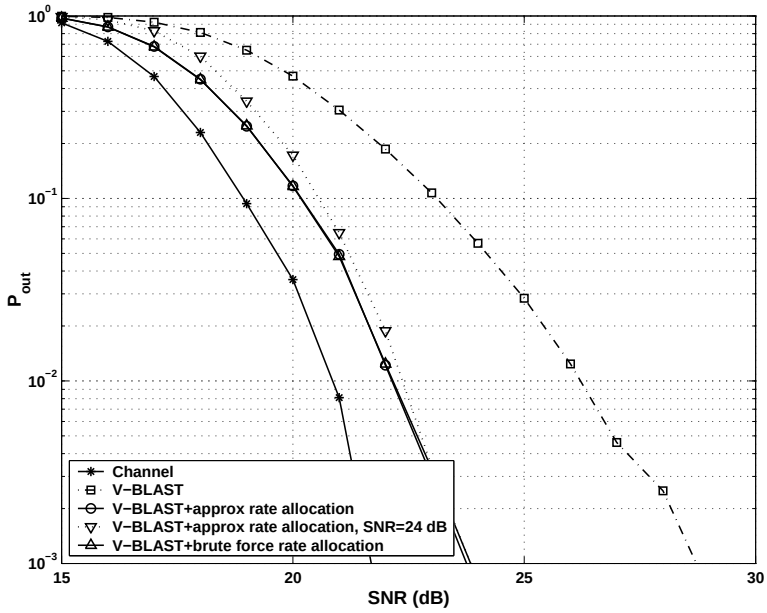
(b)

6.4 Summary

We propose a modified layered structure, where the detection order is fixed and the data rate for each layer is determined by the detection order and channel statistics. Using a Gaussian approximation of layer capacities, optimum data rate allocation algorithm and total information rate selection for maximum effective throughput are derived. Simulation results show significant performance improvement over the original V-BLAST structure and the system performance improves with increasing ergodicity within each codeword, making it promising for frequency-selective channels.

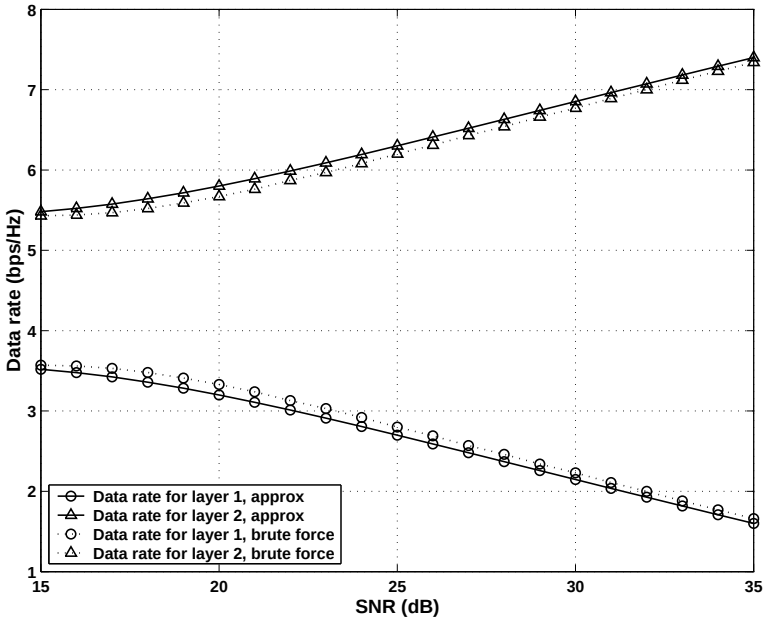


(a)

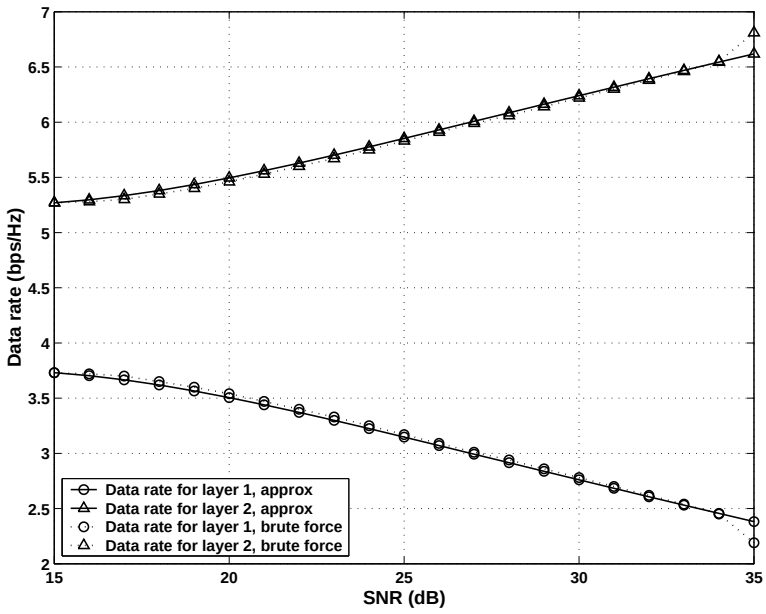


(b)

Figure 23: Outage probability of statistical rate allocation for a 2x2 layered system, (a) channel 'B', (b) Channel 'D'.

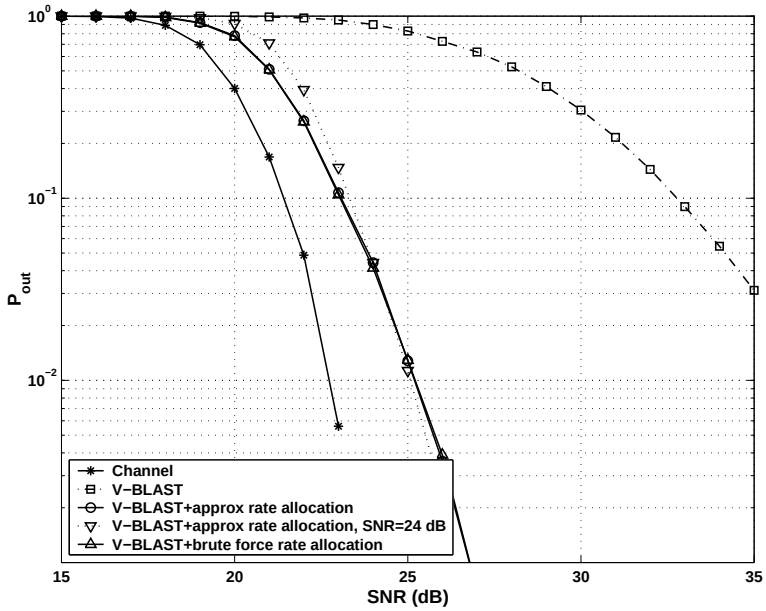


(a)

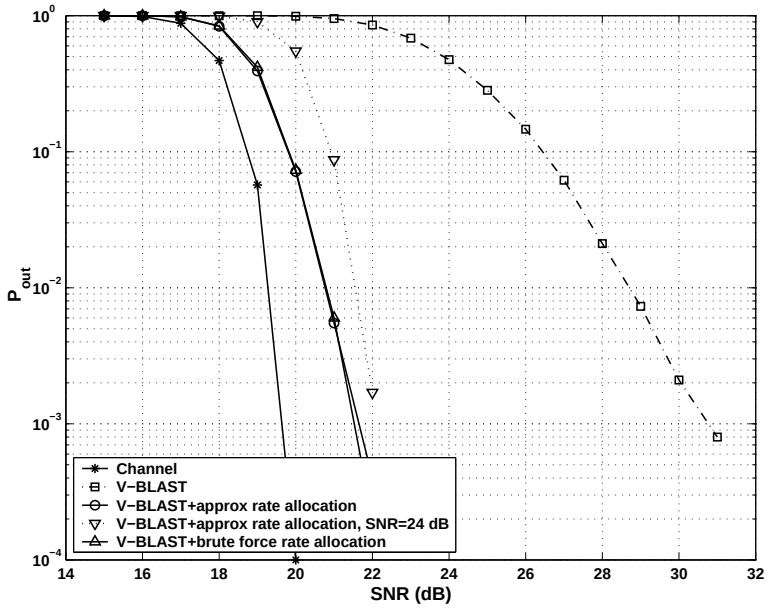


(b)

Figure 24: Statistical rate allocation for a 2x2 layered system, (a) channel 'B', (b) channel 'D'.

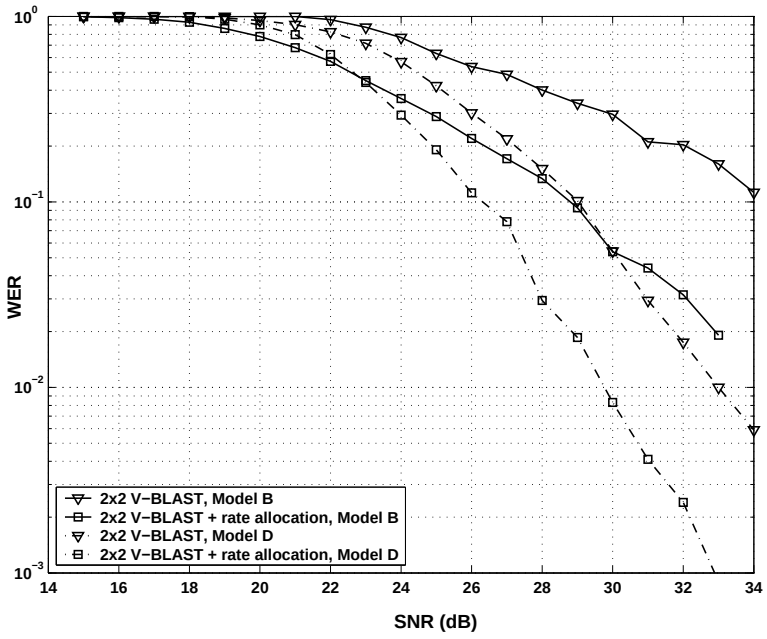


(a)

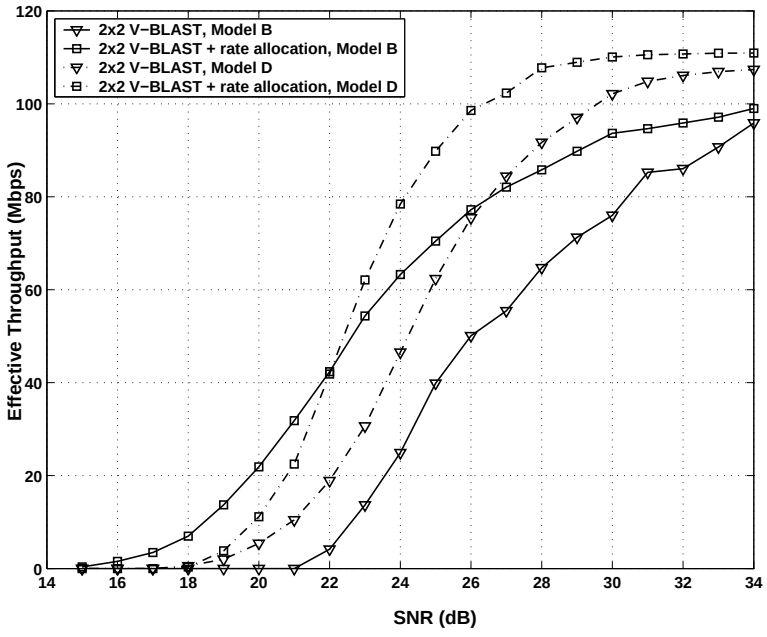


(b)

Figure 25: Outage probability for a 4x4 layered system with statistical rate allocation. (a) Channel 'B'. (b) Channel 'D'.

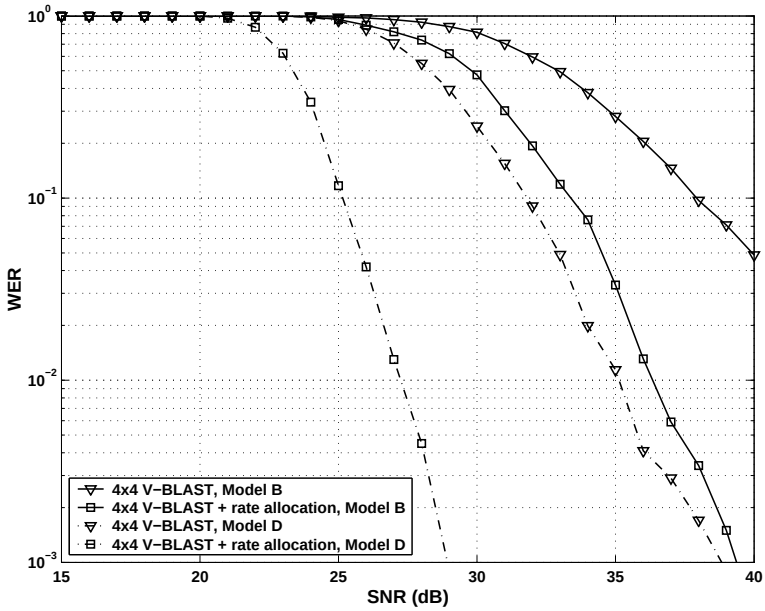


(a)

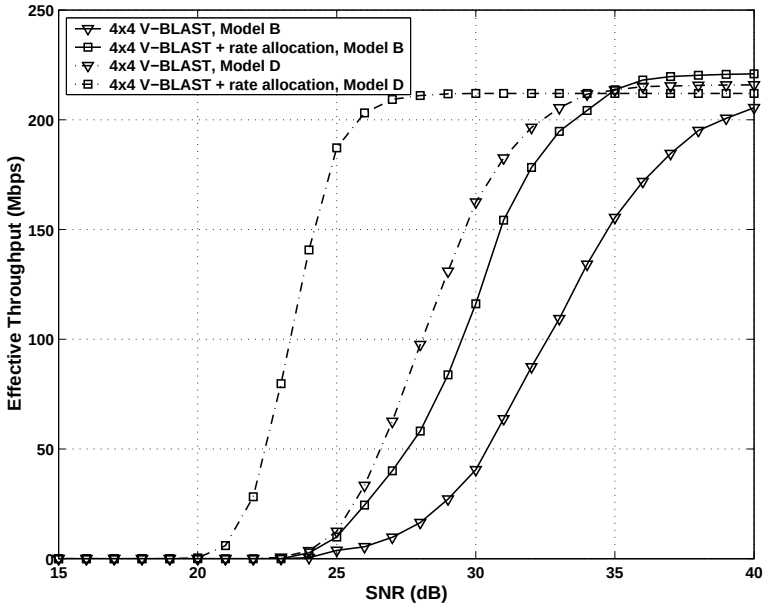


(b)

Figure 26: Simulation results for a 2x2 layered system with statistical rate allocation, 1000B packet, channel 'B' and 'D'. (a) WER. (b) Effective throughput.



(a)



(b)

Figure 27: Simulation results for a 4x4 layered system with statistical rate allocation, 1000B packet, channel 'B' and 'D'. (a) WER. (b) Effective throughput.

CHAPTER VII

CONCLUSIONS

Layered space-time structure has shown its potential as a low-complexity low-cost solution to future wireless communication systems with huge increase in spectrum efficiency and throughput. Most of the work done in this thesis is focused on performance enhancement techniques for layered space-time MIMO-OFDM systems. Our specific contributions are listed below:

- Channel estimation based on subspace tracking
- Parallel detection with predictive soft interference cancellation
- Quasi-block diagonal LDPC space-time codes
- Statistical rate allocation

First, instead of using the DFT-based channel estimation algorithm for decision-directed LS channel estimation, which has the well-known leakage problem due to channel model mismatch, we propose to estimate the dominant subspace spanned by the frequency responses of different multipaths, or equivalently, to track the optimum low-rank representation of the channel frequency response. Due to the fact that in a mobile environment the multipath delays vary slowly the subspace can be tracked by using training OFDM blocks sent at regular interval. The coefficients corresponding to different eigenvectors of the resultant subspace, which vary much faster, are updated every OFDM block based on decision-directed LS channel estimation and further refined by a filter to exploit time domain correlation. For a WER of 1%, simulation results show a performance improvement of 2 dB and 1.8 dB over the

DFT-based channel estimator, with Doppler frequency being 40 and 100 Hz, respectively. The performance improvement is explained by the significant reduction in channel estimation error floor. We further extend the channel estimation algorithm to D-BLAST OFDM systems and developed layer-wise channel estimation algorithm to update the channel estimate in a layer-wise fashion, which improves estimation accuracy of the current OFDM block for later detection stages.

Next, in Chapter 4, we study the parallel detection algorithm for group-wise space-time coding with the motivation to capitalize on the the characteristics of the space-time trellis structure for each component code. Instead of assuming that symbols from the interfering component space-time code are spatially white with zero-mean, use the partially decoded information from the Viterbi decoder to calculate the a posteriori probabilities of different trellis states of component codes given previously received signals. For component space-time codes with particular trellis structure such as those in [3], by exchanging information between the two component space-time decoder, we may be able to reduce the uncertainty in self interference, and thus improve system performance. When SNR is high, one state is much more likely than other states, and we can cancel much of the interference to the other component code from the output symbol shared by all branches from that state by subtracting the conditional mean of the interference. With soft interference cancellation rather than hard interference cancellation, we reduce the detrimental effect of error propagation due to incorrect detection of the interfering signal.

The problem with layered systems is the presence of error-propagation, and the layers to be first detected, which usually have low SNR due to loss of signal power by linear nulling, are more likely to be incorrectly decoded. In Chapter 5, we propose a space-time structure based on LDPC, which, similar to the layered space-time structure proposed in [6], can be detected with linear processing for efficient decoding. The lower-block triangular structure of the parity check matrix for the proposed

LDPC code introduces correlation between consecutive layers so that the layers can be decoded successfully with the help of information from later layers. The convergence behavior of the proposed structure is studied by means of density evolution and system simulation shows a performance improvement of 0.5 dB.

Finally, to address the detection order problem, we propose another technique called rate allocation, where the detection order is fixed and the data rate for each layer is adjusted according the detection order and corresponding channel quality, which is a promising low-complexity solution for frequency-selective channels. To avoid instantaneous rate feedback, we make use of the fact that for channels with high frequency selectivity or enough time variation, the variations of layer capacities tend to decrease, as suggested by the law of large numbers. Therefore, we can approach the open-loop channel capacity by determining the rate for each layer based on the knowledge of the channel statistics, but not of the actual realizations. Our approach is to minimize the overall layer outage probability given the total information rate. The major novelty is that we capitalize on the statistical instead of instantaneous information of the channel so that frequent rate feedback is not necessary. Using a Gaussian approximation of the layer capacities, we derive the optimum data rate allocation. The amount of backoff from the mean layer capacity is proportional to the standard deviation of the layer capacity. We then investigate how to select the total information rate to maximize the effective throughput.

APPENDIX A

PROOF OF ILL-CONDITIONING OF CHANNEL ESTIMATION WITH INCREASING RANK

Let $\mathbf{U}' = (\mathbf{U},\ \mathbf{u}_{\chi+1})$, and from $\mathbf{U}'$, we define $\mathbf{T}'[n]$ as in equation (23) by substituting $\mathbf{U}'$ for $\mathbf{U}$. We will show that

$$cond\left(\mathbf{T}'^{H}[n]\,\mathbf{T}'[n]\right) \geq cond\left(\mathbf{T}^{H}[n]\,\mathbf{T}[n]\right),$$

where $cond(.)$ means condition number of a matrix.

Proof: Let the eigenvalues of $\mathbf{T}^{H}[n]\,\mathbf{T}[n]$ be $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{\chi N_t} > 0$, From equation (23) and by exchanging columns which does not change the eigenvalues, we have

$$\mathbf{T}'[n] = (\mathbf{T}[n], \mathbf{Y}[n]),$$

where

$$\mathbf{Y}[n] = \left(\hat{\mathbf{D}}_1[n]\,\mathbf{u}_{\chi+1},\ \cdots,\ \hat{\mathbf{D}}_{N_t}[n]\,\mathbf{u}_{\chi+1}\right).$$

$$\mathbf{T}'^{H}[n]\,\mathbf{T}'[n] = \begin{pmatrix} \mathbf{T}^{H}[n]\,\mathbf{T}[n] & \mathbf{T}^{H}[n]\,\mathbf{Y}[n] \\ \mathbf{Y}^{H}[n]\,\mathbf{T}[n] & \mathbf{Y}^{H}[n]\,\mathbf{Y}[n] \end{pmatrix}$$

Let the eigenvalues of $\mathbf{T}'^{H}[n]\,\mathbf{T}'[n]$ be $\gamma'_1 \geq \gamma'_2 \geq \cdots \geq \gamma'_{(\chi+1)N_t} > 0$. By the bordering theorem for Hermitian matrices [27], we have

$$\gamma'_1 \geq \gamma_1 \geq \gamma_{\chi N_t} \geq \gamma'_{(\chi+1)N_t} > 0$$

thus

$$\begin{aligned} cond\left(\mathbf{T}'^{H}[n]\,\mathbf{T}'[n]\right) &= \frac{\gamma'_1}{\gamma'_{(\chi+1)N_t}} \\ &\geq \frac{\gamma_1}{\gamma_{\chi N_t}} \\ &= cond\left(\mathbf{T}^{H}[n]\,\mathbf{T}[n]\right). \end{aligned}$$

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, Nov-Dec 1999.

[2] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 315–335, Mar. 1998.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[4] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[5] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Systems, and Electronics*, Pisa, Italy, Oct. 1998, pp. 295–300.

[6] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, pp. 41–59, Aug. 1996.

[7] H. E. Gamal and J. A. R. Hammons, "A new approach to layered space-time coding and signal processing," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2321–2334, Sept. 2001.

[8] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

[9] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, July 2002.

[10] K. Lee and D. B. Williams, "A space-time coded transmitter diversity technique for frequency selective fading channels," in *Proc. Sensor Array and Multichannel Signal Processing Workshop 2000*, 2000, pp. 149–152.

[11] Q. Yan and R. Blum, "Optimum space-time convolutional codes," in *Proc. WCNC 2000*, vol. 3, Sept. 2000, pp. 1351–1355.

[12] S. Siwamogsatham and M. P. Fitz, "Improved high-rate space-time codes via orthogonality and set partitioning," in *Proc. WCNC 2002*, Mar. 2002, pp. 264–270.

[13] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data-rate wireless communication over wideband channels," in *Proc. 48th IEEE Vehicular Technology Conf.*, vol. 3, May 1998, pp. 2232–2236.

[14] H. Bolcskei and A. Paulraj, "Space-frequency coded broadband OFDM systems," in *Proc. WCNC'02*, vol. 1, May 2000, pp. 1–6.

[15] Y. Gong and K. Letaief, "An efficient space-frequency coded wideband OFDM system for wireless communications," in *Proc. ICC 2002*, May 2002, pp. 475–479.

[16] N. Boubaker, K. Letaief, and R. Murch, "A layered space-time coded wideband ofdm architecture for dispersive wireless links," in *Proc. 6th IEEE Symposium on Computers and Communications, 2001*, July 2001, pp. 518–523.

[17] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1121–1128, May 1999.

[18] R. Blum, Y. G. Li, J. Winters, and Y. Qing, "Improved space-time coding for MIMO-OFDM wireless communications," *IEEE Trans. Commun.*, vol. 49, pp. 1873–1878, Nov. 2001.

[19] C. B. Papadias and G. J. Foschini, "Capacity-approaching space-time codes for systems employing four transmitter antennas," *IEEE Trans. Inform. Theory*, vol. 49, pp. 726–733, Mar. 2003.

[20] Y. G. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 461–471, Mar. 1999.

[21] Y. G. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 67–75, Jan. 2002.

[22] J.-J. van de Beek, O. S. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. 45th IEEE Vehicular Technology Conf.*, Chicago, IL, July 1995, pp. 815–819.

[23] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, pp. 467–479, Mar. 2001.

[24] R. Chen and K. B. Letaief, "Channel estimation for space time coded OFDM systems in non-sample-spaced multipath channels," in *Proc. WCNC'02*, Orlando, FL, Mar. 2002, pp. 61–66.

[25] T. A. Thomas and F. W. Vook, "Broadband MIMO-OFDM channel estimation via near-maximum likelihood time of arrival estimation," in *Proc. ICASSP'02*, vol. 3, Orlando, Fl, May 2002, pp. 2569–2572.

[26] Y. G. Li, L. J. Cimini, and N. R. Sollegberger, "Robust channels estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 902–915, July 1998.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, 1985.

[28] O. Edfors, M. Sandell, J.-J. van de Beek, *et al.*, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, July 1998.

[29] I. Barhumi, G. Leus, and M. Moonen, "Optimal training sequences for channel estimation in MIMO OFDM systems in mobile wireless channels," in *Proc. 2002 International Zurich Seminar on Broadband Communications*, Zurich, Switzerland, Feb. 2002, pp. 44–1–44–6.

[30] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 387–392, Apr. 1985.

[31] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty-III: The dimension of the space of essentially time- and band-limted signals," *Bell Syst. Tech. J.*, vol. 41, pp. 1295–1336, July 1962.

[32] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Processing*, vol. 44, pp. 2932–2947, Dec. 1996.

[33] Q. Yan and R. Blum, "Optimum space-time convolutional codes," in *WCNC 2000*, vol. 3, 2000, pp. 1351–1355.

[34] R. A. Monzingo and T. W. Miller, *Introduction To Adaptive Arrays*. John Wileys & Sons, 1980.

[35] W. E. Ryan and S. G. Wilson, "Two classes of convolutional codes over gf(q) for q-ary orthogonal signaling," *IEEE Trans. Commun.*, vol. 39, pp. 30–40, Jan. 1991.

[36] Y. G. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for wireless communications: signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, vol. 50, pp. 1471–1477, Sept. 2002.

[37] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[38] W. J. Choi, K. W. Cheong, and J. M. Cioffi, "Iterative soft interference cancellation for multiple antenna systems," in *Proc. IEEE WCNC'00*, vol. 1, Chicago, IL, Sept. 2000, pp. 304–309.

[39] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Commun.*, vol. 48, pp. 1347–1359, Aug. 2000.

[40] S. Sun, T. T. Tjhung, and P. H. W. Fung, "Soft decision-based iterative interference cancellation(IIC) in group-wise STBC (G-STBC) MIMO systems," in *Proc. 57th IEEE Vehicular Technology Conf.*, vol. 2, Apr. 2003, pp. 984–988.

[41] X. Zhu and R. D. Murch, "MIMO-DFE based blast over frequency selective channels," in *Proc. IEEE GLOBECOM'01*, vol. 1, San Antonio, TX, Nov. 2001, pp. 499–503.

[42] S. Baro, G. Bauch, A. Pavlic, and A. Semmler, "Improving BLAST performance using space-time block codes and turbo decoding," in *Proc. IEEE GLOBECOM'00*, vol. 2, San Francisco, CA, Nov. 2000, pp. 1067–1071.

[43] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo codes," in *ICC*, 1993, pp. 1064–1070.

[44] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399–431, 1999.

[45] Y. Kou, S. Lin, and M. P. C. Fossorier, "Low-density parity-check codes based on fininte geometries: a rediscovery and new results," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2711–2736, 2001.

[46] R. G. Gallager, *Low-Density Parity-Check Codes.* Cambridge, MA : MIT Press, 1963.

[47] F. Kschischang, B. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, 2001.

[48] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, 1974.

[49] S. Y. Chung, G. D. F. Jr., T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 db of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, pp. 58–60, 2001.

[50] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufmann, 1988.

[51] T. Richardson and R. Unbanke, "The renaissance of Gallager's low-density parity-check codes," *IEEE Commun. Magazine*, pp. 126–131, 2003.

[52] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman, "Analysis of low density codes and improved designs using irregular graphs," in *30th Annu. ACM Symp. Theory of Computing*, 1998, pp. 249–258.

[53] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, 2001.

[54] S. Y. Chung, T. J. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, 2001.

[55] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, 2001.

[56] T. J. Richardson, M. A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, 2001.

[57] S. t. Brink, "Iterative decoding trajectories of parallel concatenated codes," in *3rd IEEE ITG Conf. Source and Channel Coding*, Munich, Germany, 2000.

[58] ——, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, pp. 1117–1118, 1999.

[59] S. t. Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, pp. 670–678, 2004.

[60] P. Meshkat and H. Jafarkhani, "Space-time low-density parity-check codes," in *36th Asilomar Conference on Signals, Systems and Computers*, 2002, pp. 1117–1121.

[61] J. Hou, P. H. Siegel, and L. B. Milstein, "Performance analysis and code optimization of low density parity-check codes on Rayleigh fading channels," *IEEE J. Selected Areas Comm.*, vol. 19, pp. 924–934, 2001.

[62] B. Lu, X. Wang, and K. R. Narayanan, "LDPC-based space-time coded OFDM systems over correlated fading channels: performance analysis and receiver design," *IEEE Trans. Commun.*, vol. 50, pp. 74–88, 2002.

[63] B. Lu, G. Yue, and X. Wang, "Performance analysis and design optimization of LDPC-coded MIMO OFDM systems," *IEEE Trans. Signal Processing*, vol. 52, pp. 348–361, 2004.

[64] V. Marnnoni, D. Declercq, and G. Gelle, "Optimized irregular low-density parity-check codes for multicarrier modulations over frequency selective channels," *EURASIP Journal on Applied Signal Processing*, pp. 1546–1556, 2004.

[65] B. M. Hochwald and S. t. Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Wireless Commun.*, vol. 51, pp. 389–399, 2003.

[66] S. Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, Feb. 2001.

[67] X. Y. Hu, E. Eleftheriou, D. M. Arnold, and A. Dholakia, "Efficient implementations of the sum-product algorithm for decoding ldpc codes," *GLOBECOM 2001*, vol. 2, pp. 25–29, Nov. 2001.

[68] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[69] J. Hou, P. H. Siegel, and L. B. Milstein, "Performance analysis and code optimization of low density parity-check codes on Rayleight fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 924–934, May 2001.

[70] B. Lu, G. Yue, and X. Wang, "Performance analysis and design optimization of LDPC-coded MIMO OFDM systems," *IEEE Trans. Signal Processing*, vol. 52, pp. 348–361, Feb. 2004.

[71] L. Schumacher, "Matlab implementation of the IEEE 802.11 HTSG channel model special committee proposal." [Online]. Available: http://www.info.fundp.ac.be/∼lsc/Research /IEEE_80211_HTSG_CMSC/distribution_terms.html

[72] X. Hu. Source code for progressive edge growth parity-check matrix construction. [Online]. Available: http://www.inference.phy.cam.ac.uk/ mackay/PEG_ECC.html

[73] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST : An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Systems, Electronics*, Sept. 1998, pp. 295–300.

[74] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *IEEE Electr. Lett.*, vol. 35, pp. 14–16, Jan. 1999.

[75] S. H. Nam, O. Shin, and K. B. Lee, "Transmit power allocation for a modified v-blast system," *IEEE Trans. Commun.*, vol. 7, pp. 1074–1078, July 2004.

[76] S. Catreux, P. f. Driessen, and L. J. Greenstein, "Data throughput using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environment," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 226–234, Apr. 2002.

[77] H. Zhuang, L. Dai, S. Zhou, and Y. Yao, "Low complexity per-antenna rate and power control approach for closed-loop V-BLAST," *IEEE Trans. Commun.*, vol. 11, pp. 1783–1787, Nov. 2003.

[78] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *IEEE Asilomar Conference on Signals, Systems, and Computers*, Nov. 1997, pp. 1405–1409.

[79] S. T. Chung, A. Lozano, and H. Huang, "Approaching eigenmode BLAST channel capacity using V-BLAST with rate and power feedback," in *Proc. VTC 2001*, Oct. 2001, pp. 915–919.

[80] A. Lozano, "Capacity-approaching rate function for layered multiantenna architectures," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 616–620, July 2003.

[81] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. ICC 2002*, Apr. 2002, pp. 406–410.

[82] M. A. Kamath, B. L. Hughes, and X. Yu, "Gaussian approximations for the capacity of MIMO Rayleigh fading channels," in *IEEE Asilomar Conference on Signals, Systems, and Computers*, Nov. 2002, pp. 614–618.

[83] S. Verdu and S. Shamai, "Spectral efficiency of cdma with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.

[84] P. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Commun.*, vol. 48, pp. 1245–1248, Aug. 2000.

[85] J. G. Proakis, *Digital Communications*. McGraw Hill, 1989.

INDEX

VITA

Jianxuan Du was born in Xi'an, Shaanxi Province, P. R. China. He obtained his B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China, in 1998 and 2001, respectively, both in Electrical Engineering. Since the fall of 2001, he has been a Graduate Research Assistant at Information Transmission and Processing Laboratory at Georgia Institute of Technology. From Jan 2004 to July 2004, he worked as an intern with Mitsubishi Electric Research Laboratories (MERL) on the IEEE 802.11n standard project. His research interests include signal detection, channel estimation and channel coding for wireless channels, with emphasis on MIMO-OFDM systems.